Seeing Photons in Color

SIZHUO MA, University of Wisconsin-Madison, USA and Snap Inc., USA
VARUN SUNDAR, University of Wisconsin-Madison, USA
PAUL MOS, CLAUDIO BRUSCHINI, and EDOARDO CHARBON, EPFL, Switzerland
MOHIT GUPTA, University of Wisconsin-Madison, USA

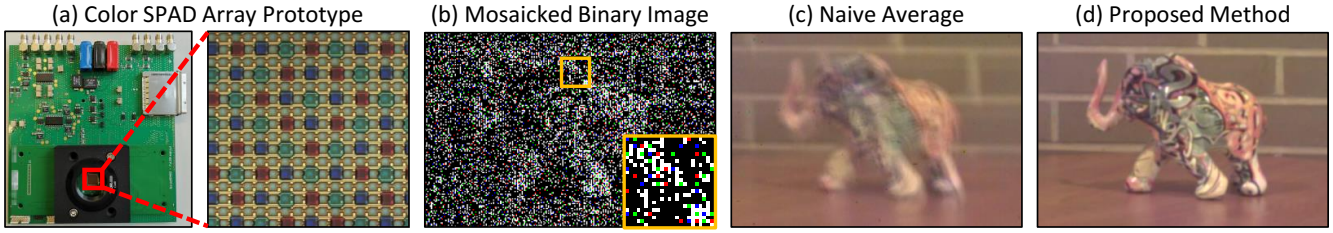

Fig. 1. **Seeing photons in color. (a)** In this paper, we propose a photon-counting color imaging system. We design and fabricate a pseudorandom RGBW color filter array (CFA) for a color SPAD array, which captures mosaicked binary images at 496×254, up to 96.8kfps. **(b)** A single mosaicked binary image is highly noisy and does not contain sufficient information for reconstructing colors. **(c)** One naive idea is to take the average of a sequence of mosaicked quanta images and them perform demosaicking. However, this approach results in blur when there is considerable motion between the camera and the scene—especially in low-light scenes which require longer exposure times. **(d)** We propose an algorithm that aligns the quanta frames, jointly demosaics and merges them into a single intensity image and applies spatial denoising to generate a clean, blur-free image. **Zoom in for details.**

Megapixel single-photon avalanche diode (SPAD) arrays have been developed recently, opening up the possibility of deploying SPADs as general-purpose passive cameras for photography and computer vision. However, most previous work on SPADs has been limited to monochrome imaging. We propose a computational photography technique that reconstructs high-quality color images from mosaicked binary frames captured by a SPAD array, even for high-dyanamic-range (HDR) scenes with complex and rapid motion. Inspired by conventional burst photography approaches, we design algorithms that jointly denoise and demosaick single-photon image sequences. Based on the observation that motion effectively increases the color sample rate, we design a blue-noise pseudorandom RGBW color filter array for SPADs, which is tailored for imaging dark, dynamic scenes. Results on simulated data, as well as real data captured with a fabricated color SPAD hardware prototype shows that the proposed method can reconstruct high-quality images with minimal color artifacts even for challenging low-light, HDR and fast-moving scenes. We hope that this paper, by adding color to computational single-photon imaging, spurs rapid adoption of SPADs for real-world passive imaging applications.

CCS Concepts: • **Computing methodologies** → **Computational photography**.

Authors' addresses: Sizhuo Ma, sizhuoma@cs.wisc.edu, University of Wisconsin-Madison, USA and Snap Inc., USA; Varun Sundar, vsundar4@wisc.edu, University of Wisconsin-Madison, USA; Paul Mos, paul.mos@epfl.ch; Claudio Bruschini, claudio.bruschini@epfl.ch; Edoardo Charbon, edoardo.charbon@epfl.ch, EPFL, Switzerland; Mohit Gupta, mohitg@cs.wisc.edu, University of Wisconsin-Madison, USA.

0730-0301/2023/8-ART $15.00
https://doi.org/10.1145/3592438

Additional Key Words and Phrases: Single-photon camera, single-photon avalanche diode, quanta image sensor, burst photography, color filter array, demosaicking, high dynamic range, high-speed imaging, low-light imaging

ACM Reference Format:
Sizhuo Ma, Varun Sundar, Paul Mos, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. 2023. Seeing Photons in Color. *ACM Trans. Graph.* 42, 4 (August 2023), 16 pages. https://doi.org/10.1145/3592438

1 INTRODUCTION

Single-photon avalanche diodes (SPADs) are an emerging class of image sensors that can record individual photons with precise timing. Because of this exciting capability, they have been used for various active imaging applications such as LiDAR [Li et al. 2017], non-line-of-sight imaging (NLOS) [Buttafava et al. 2015] and fluorescence lifetime imaging (FLIM) [Bruschini et al. 2019]. Although SPADs were limited to single-pixel or low-resolution (e.g., 32×32) form-factors in these applications, the last few years have witnessed a single-photon revolution, culminating in the development of high-resolution SPAD arrays for the first time (1/4 MPixel [Ulku et al. 2019], 1 MPixel [Morimoto et al. 2020], 3.2 MPixel [Morimoto et al. 2021]). These arrays capture binary frames at high speeds (up to 100kfps), with negligible read noise. Such unique properties make them a potential alternative to conventional CMOS sensors for *passive imaging*, especially for high dynamic range (HDR), fast-moving scenes. This opportunity to deploy them as general-purpose passive cameras opens up a considerably wider range of applications for SPADs (beyond just scientific and 3D imaging), including machine and robot vision, and consumer photography.

So far, most research on SPAD arrays has focused on monochrome imaging, with surprisingly little attention paid to color. There is no doubt that color is important, often critical, not only for capturing captivating photographs, but also for machine vision tasks such as

detection and recognition [Khan et al. 2012]. This raises a natural question: How do we add color to single-photon imaging?

One approach is to fabricate color filter arrays (CFAs) on a SPAD array, much like conventional cameras. The resulting color SPAD array will capture *color mosaicked* binary images. However, unlike conventional images, single binary (1-bit) frames do not contain sufficient information to be demosaicked directly with conventional demosaicking algorithms. One could capture a *series* of mosaicked binary images over time, and sum them to create a mosaicked intensity image with sufficient signal and dynamic range, and then apply demosaicking algorithms to get an RGB image. However, if there is motion, the photons emitted by a given scene point are dispersed over multiple pixels across the binary image sequence, resulting in blur. The blur is especially severe when capturing dark scenes, where longer sequence (equivalent to longer exposure for conventional cameras) are needed to gather enough light.

Burst photography with a color SPAD array. The challenge for dark, dynamic scenes can be summarized as a trade-off between blur and noise: a short sequence leads to less blur but noisier images due to photon noise, while longer sequences give cleaner images, albeit with more significant motion blur. This trade-off can be mitigated by burst photography [Hasinoff et al. 2016; Wronski et al. 2019], which divides a long exposure into a number of frames with short exposures, computes and compensates for the motion between frames, and then merges them into a blur-free, low-noise image. Similar approaches have been developed for monochrome SPADs [Gyongy et al. 2018; Iwabuchi et al. 2021; Ma et al. 2020; Seets et al. 2021], jots [Chi et al. 2020] and spike cameras [Zhao et al. 2021].

The key challenge that prevents applying conventional burst photography directly to mosaicked binary frames is that brightness constancy does not hold. It is difficult, if not impossible, to estimate the motion between mosaicked binary frame directly because a scene point may "move", for example, from a green filter to a red filter across the sequence. To address this challenge, we develop the *first burst photography algorithm for color SPADs*, which processes the raw mosaicked data at the granularity of individual photons. The proposed approach can be considered a universal demosaicking algorithm for frame-based SPAD cameras with *any* CFA consisting of R, G, B and W pixels, with regular or pseudorandom layouts.

Designing color filter arrays for SPADs. What is the best CFA for color SPADs? In addition to red, green and blue pixels, white pixels that are sensitive to all three colors absorb more light and give superior image quality in the dark. Such RGBW patterns are a good candidate for low-light imaging with SPADs. However, there is one known challenge for RGBW arrays: Due to the sparseness of color filters, the reconstructed images are often subject to aliasing artifacts such as moiré color banding or false colors. Our key observation is that, when the scene or the camera is moving, the effective color sampling rate of RGBW arrays is increased, *especially due to the high temporal sampling rate of SPADs*, which mitigates the color artifacts while benefiting from the high SNR of W pixels. Inspired by previous work on RGB patterns, we design a blue-noise pseudorandom RGBW CFA for color SPADs, which further boosts the quality of low-light burst photography. We fabricate the CFA with photolithography on our SPAD array, resulting in, to our knowledge, the first unconventional RGBW CFA implemented on SPADs. Our color SPAD prototype, for the first time, offers low-level access to mosaicked single-photon binary frames, as opposed to only time-integrated photon counts [Morimoto et al. 2021].

Scope and limitations. The proposed methods are applicable to a generalized class of single-photon cameras (SPC) including not only SPADs, but also *jots* [Fossum 2005], which have smaller pixel sizes, but lower frame rates than SPADs. We show via simulations that jots-based color burst photography complements SPADs in scenarios where high-frequency spatial details need to be recovered and only low-speed motion is involved.

Although we show promising image reconstructions with simulated and real color SPAD arrays, the proposed single-photon imaging system is not ready to directly compete with current CMOS image sensors (*e.g.*, on smartphones). Specifically, image sensors on mobile devices have strict constraints on pixel pitch, power consumption, processing efficiency, *etc*, which current SPAD technology cannot meet yet. We envision that SPAD hardware and processing algorithms will continue to mature in the coming years, and the techniques introduced in this work may lead to increased interest in, and accelerate the future development of, passive SPAD imaging.

2 RELATED WORK

Single-photon cameras. There are two main families of single-photon cameras: SPADs and jots, often referred to as SPAD-QIS and CIS-QIS (quanta image sensor) as well, respectively. SPADs record the arrival of photons by amplifying the weak signal of single incident photons via avalanche multiplication [Zappa et al. 2007]. As a result, extremely high frame rates are achieved for even large format SPAD arrays (97.7kfps for 1/4MPixel [Ulku et al. 2019] and 24kfps for 1MPixel [Morimoto et al. 2020]) with virtually no read noise. Recent developments have focused on increasing spatial resolution (3.2MPixel [Morimoto et al. 2021]) and HDR [Ogi et al. 2021; Ota et al. 2022]. Jots avoid photon avalanche and achieve high sensitivity by using an active pixel architecture with low capacitance and high conversion gain [Fossum 2005]. As a result, jots have a smaller pixel pitch, higher quantum efficiency, but capture binary images at a lower frame rate (1040fps [Ma et al. 2017]). The proposed approach can be applied to both kinds of sensors as they share the same mathematical imaging model, described in Sec. 3.

Burst photography for conventional cameras. Conventional burst photography takes a series of underexposed intensity images and combines them into a single high-quality image. Motion compensation can be done by an explicit align-and-merge approach [Hasinoff et al. 2016; Liba et al. 2019; Liu et al. 2014; Wronski et al. 2019], or jointly through optimization [Heide et al. 2016, 2014]. Recently, deep neural networks have been developed for burst denoising, which either combine frames directly [Bhat et al. 2021; Dudhane et al. 2022; Godard et al. 2018; Liang et al. 2020], or predict kernels that re-weight and merge images [Mildenhall et al. 2018; Xia et al. 2019]. Learning-based burst photography [Chen et al. 2019, 2018; Dong et al. 2022; Jiang and Zheng 2019; Karadeniz et al. 2021; Kokkinos and Lefkimmiatis 2019] operates directly on raw images and shows

outstanding performance for extreme low-light scenes. Neural radiance fields have been applied to burst photography [Mildenhall et al. 2021; Pearl et al. 2022], which handle large motion and high noise level well. While this work is inspired by the classical two-step approach to demonstrate the feasibility and benefits of color burst photography with SPADs, neural network-based approaches are a promising future direction, as discussed in Sec. 8.

Image reconstruction for single-photon cameras. Prior work has analyzed the statistics of single-photon images for static scenes [Antolovic et al. 2016; Yang et al. 2012], and developed methods for reconstructing intensity images via standard denoising techniques such as total variation and BM3D [Chan et al. 2016; Gnanasambandam et al. 2019], or using end-to-end neural networks [Chandramouli et al. 2019; Choi et al. 2018]. Fossum [2013] first suggested that sequential binary frames should be shifted to compensate for relative motion, which is later implemented by assuming simple global motion models [Gyongy et al. 2018; Iwabuchi et al. 2021, 2019; Seets et al. 2021]. Chi et al. [2020] leverages student-teacher learning to achieve both denoising and deblurring for a short quanta image sequence (8 frames). Inspired by conventional burst photography, [Ma et al. 2020] makes less restrictive assumption on the motion (patch-wise 2D translation, smooth in time) and can deal with non-rigid scene motion over a long sequence (1000-10000 binary frames). In this paper, we propose the first image reconstruction algorithm for color SPADs *in the presence of scene and camera motion.*

Color imaging for single-photon cameras. Color imaging with single-photon cameras have been demonstrated with active lighting [Griffiths et al. 2019; Ren et al. 2018]. Gnanasambandam et al. [2019] developed the first megapixel jot camera with a Bayer RGB CFA pattern, and demonstrated promising results for joint demosaicking-denoising [Elgendy et al. 2021] and image classification [Gnanasambandam and Chan 2020]. Elgendy et al. [2020] studied the spatio-spectral design of CFA for jot cameras using an optimization-based framework. Shah et al. [2020] built the first color SPAD array using RGB plasmonic metasurface mosaic filters, with a resolution of 64×64. Morimoto et al. [2021] demonstrated the first multi-megapixel color SPAD array with an RGB Bayer pattern, but does not consider motion. In contrast, this paper analyzes and implements unconventional RGBW color filters on a large-format SPAD array, and develops computational approaches for performing motion-compensation and demosaicking on single-photon frame sequences captured by a color SPAD array.

3 SINGLE-PHOTON IMAGING MODEL

In this section, we describe the imaging model for single-photon cameras in the context of color photography. Detailed explanation and analysis about monochrome photography can be found in previous work [Antolovic et al. 2016; Ma et al. 2020; Yang et al. 2012].

Consider a single-photon sensor, e.g., a SPAD pixel array. Let the light incident on a pixel be given by $\phi(\lambda)$, the *spectral photon flux* which describes the average number of incident photons per second as a function of wavelength λ.[1] When a photon of wavelength λ hits a SPAD pixel, the probability of triggering an avalanche is called photon detection efficiency (PDE) $\eta(\lambda)$. Next, suppose the sensor is covered with a color filter array so that different pixels have filters of different colors, which transmit only a selected wavelength band of light and thus, have different $\eta(\lambda)$. Then, the average number of photon counts per second is:

$$\rho_c = \int \phi(\lambda)\eta_c(\lambda)\mathrm{d}\lambda, \qquad c = R, G, B, W\,. \tag{1}$$

We call ρ_c *color intensity* of channel c, which is the quantity we want to estimate at each pixel.

Note that ρ_c only represents the average counts of photoelectrons for channel c. The *actual* number of excited photoelectrons Z_c during an exposure time of τ seconds is modeled as a Poisson random variable:

$$P\{Z_c = k\} = \frac{(\rho_c\tau)^k e^{-\rho_c\tau}}{k!}\,, \tag{2}$$

A single-photon sensor pixel records at most one photon during an exposure, returning a binary value B such that $B = 1$ if the pixel detects one or more photons, and $B = 0$ otherwise. B is therefore a random variable with Bernoulli distribution (for simplicity, we drop the subscript c in the following):

$$\begin{aligned} P\{B = 0\} &= e^{-(\rho\tau + r_d\tau)}\,, \\ P\{B = 1\} &= 1 - e^{-(\rho\tau + r_d\tau)}\,, \end{aligned} \tag{3}$$

where r_d is the dark count rate (DCR), the rate of spurious counts unrelated to photon arrivals. We call $B(x, y)$ a *mosaicked quanta image*, where (x, y) represents the 2D pixel location.

If the scene and camera are static, the color intensity ρ can be estimated by capturing a sequence of mosaicked binary frames, and then simply adding them together to form a *sum image* $S(x, y)$:

$$S(x, y) = \sum_{t=1}^{n} B_t(x, y)\,, \tag{4}$$

where $B_t(x, y)$ is the binary frame at time t, and n is the number of frames. $S(x, y)$ is the total photon counts at (x, y) over the entire binary image sequence. Since each binary frame is independent, the expected value of the sum image is the product of the number of frames n, and the expected value of the Bernoulli variable B:

$$E[S(x, y)] = n\,E[B(x, y)] = n\left(1 - e^{-(\rho\tau + r_d\tau)}\right). \tag{5}$$

The maximum likelihood estimate (MLE) of the color intensity ρ is given by [Antolovic et al. 2016]:

$$\hat{\rho}(x, y) = -\ln(1 - S(x, y)/n)/\tau - r_d(x, y)\,. \tag{6}$$

Color intensity estimation under motion: So far we assumed the scene and camera are static such that ρ is constant during the capture. When the scene or the camera are moving, the motion needs to be estimated so that ρ remains constant after compensating for motion. This is particularly challenging in the presence of a color filter array, due to which the photon counts ρ depends not only on scene brightness (ϕ), *but also filter spectral properties* (η). In the next section, we describe techniques for generating high-quality, blur-free linear RGB image from the captured mosaicked quanta image sequences, in the presence of motion.

[1] Strictly speaking, since λ is a continuous quantity, $\phi(\lambda)$ describes the spectral density of the incident photon flux.

4 MOTION AND MOSAICKING IN QUANTA IMAGES

In this section, we propose techniques for reconstructing RGB images from a sequence of mosaicked quanta frames. Broadly, the approach consists of two components: First, we adopt a hierarchical approach to compute motion between a reference frame and every other mosaicked 1-bit frame, which is challenging due to extreme quantization (1-bit), low signal-to-noise ratio (SNR) and bandpass filtering by the CFA. Second, we propose a novel *joint demosaicking and merging algorithm* based on the observation that quanta images, due to their unique characteristics as described above, are not amenable to sequential demosaicking and merging. At the end of the section, we propose techniques to address other challenges, including a pre-processing hot pixel correction step and a chrominance-focused denoising technique.

4.1 Estimating Motion between Mosaicked Quanta Images

Each captured mosaicked quanta image contains only 1-bit information per pixel, which is not sufficient for directly estimating the motion between them. One potential solution is to divide the entire sequence into temporal blocks, compute the sum of photon counts at each pixel for each block to get a multi-bit image [Ma et al. 2020] which have sufficient SNR, and then estimating motion between these multi-bit images. However, these multi-bit images are still mosaicked (we define as *mosaicked block-sum images*), and not ready to be directly aligned. Fig. 2 shows a minimal example, with a scene consisting of a yellow cube. When the cube moves by one pixel, it shifts to a different set of color filters on the sensor and the mosaicked images look completely different, i.e. brightness constancy is violated, making it challenging to estimate motion.

To address this conundrum, we propose converting mosaicked block-sum images to grayscale images before matching. Since the densely sampled W channel (75% W pixels, see Sec. 5) carries sufficient information for alignment, we directly interpolate the W pixels [Bornemann and März 2007] to get a full-resolution grayscale image. This approach can be extended to a general class of CFAs. **Please see the supplementary report for details.**

We then apply a hierarchical patch-matching algorithm to compute the block-level motion, which is then linearly interpolated to get the fine-grained frame-level motion [Ma et al. 2020]. The block size is chosen between 100 and 1000 quanta images, depending on the camera capture frame rate, scene light level and motion speed. Fig. 3 visually summarizes the motion estimation step.

4.2 Joint Demosaicking and Merging

After alignment, we merge the binary frames into a high-SNR, low-blur color image. The merging problem is challenging since the raw frames are heavily quantized and mosaicked, which necessitates solving both the merging and demosaicking problems.

One straightforward approach is to demosaic the individual frames, followed by merging. However, with single-photon cameras, it is extremely difficult to demosaic individual binary frames since they lack reliable color information. Another idea is to demosaic the block-sum images, and then merge them. In this case, the merged image still contains motion blur within each block, which does not fully utilize the high frame rate of single-photon cameras. In summary, neither *demosaic first and then merge* nor *merge first and then demosaic* gives satisfactory results.

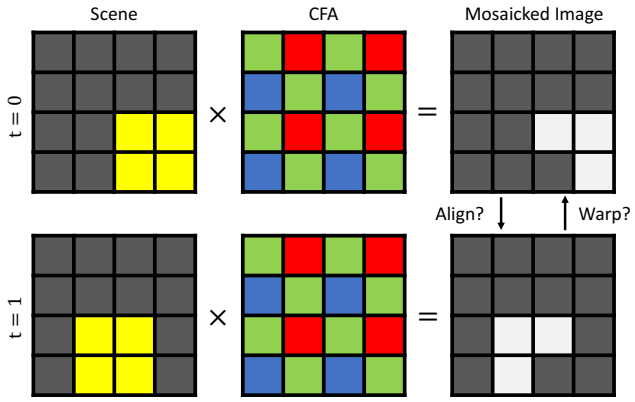

Fig. 2. **Can we align/warp mosaicked images directly?** When a yellow cube moves by one pixel, the same scene point moves to a pixel with a different color filter, which results in a completely different mosaicked image. Therefore, it is impossible to align or warp mosaicked images directly.

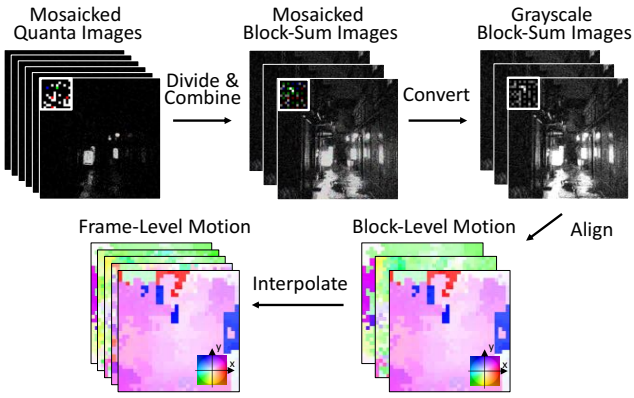

Fig. 3. **Motion estimation algorithm.** A sequence of mosaicked quanta images is divided into temporal blocks and added up to mosaicked block-sum images, which are then converted to grayscale block-sum images (as shown in Fig. 2). Block-level motion is estimated between grayscale block-sum images using a hierarchical patch-matching algorithm, which is then linearly interpolated to get frame-level motion.

Joint demosaicking and merging: Since neither of the aforementioned sequential approaches are adequate, we propose a joint demosaic-merge technique. Inspired by [Wronski et al. 2019], we treat pixels in the mosaicked binary frames as 1-bit color samples, which are warped to the reference frame according to the measured sub-pixel frame-level motion (super-resolution can be enabled by choosing a pixel grid of higher resolution). We reconstruct each pixel on the grid from samples within a spatial neighborhood, using an anisotropic Gaussian kernel [Takeda et al. 2007]. To estimate the kernel at each pixel location, we create a reference image by warping and combining the grayscale block-sum images generated in the alignment step using a Wiener filter [Hasinoff et al. 2016].

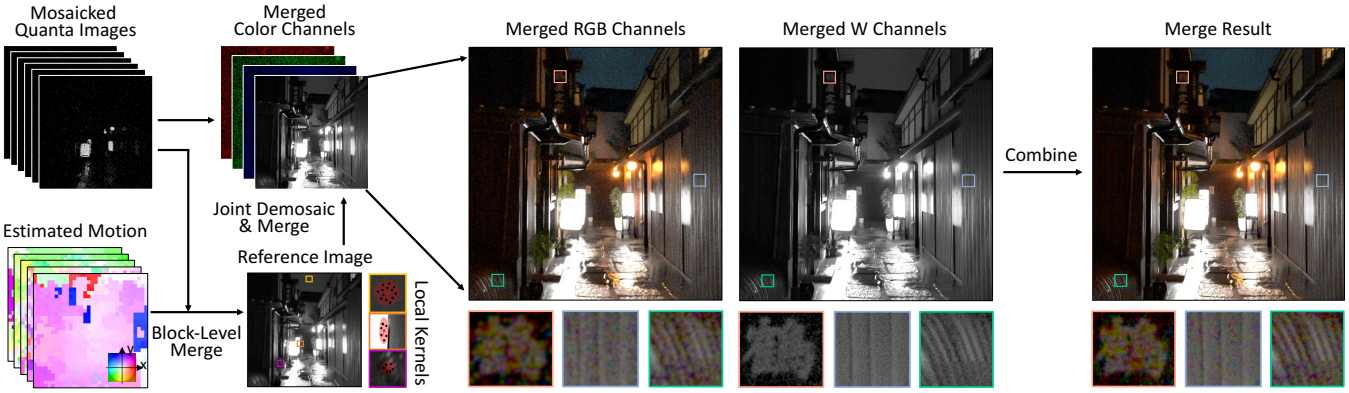

Fig. 4. **Merge algorithm.** We first use the estimated block-wise motion to warp the grayscale block-sum images to obtain a reference image, which is then used to guide the joint demosaicking and merging of mosaicked quanta images into full-resolution R, G, B, and W channels respectively. The four channels are then combined into an RGB image, where the luminance comes from the W channel and the chromaticity comes from the R, G and B channels. Since the W channel has higher SNR and spatial resolution, the merged image has improved spatial details and reduced noise as compared to the raw RGB channels.

The reference image also enables robust merging of the binary color samples, where we design a robust weighting function based on the binomial statistics. **More details about the algorithm can be found in the supplementary technical report.**

Combining R, G, B and W measurements. The proposed joint demosaicking and merging algorithm is applied to each color channel separately. When an RGBW CFA is used, the algorithm produces four full-resolution images, one each for R, G, B and W channels. Although it is possible to simply ignore the W channel and output the R, G and B channels as the final image, the W channel provides a more accurate estimate of the luminance of the image. This is because W pixels receive more light and achieve higher SNR in low-light environments, especially for CFAs with high fraction of W pixels. Thus, it is important to use the additional W channel while reconstructing the final RGB image. Formally, the W channel offers a per-pixel linear constraint on the color values:

$$w_R R + w_G G + w_B B = W\,, \tag{7}$$

where (w_R, w_G, w_B) are the color transform coefficients that relate the R, G, B and W channels and are determined by calibrating the spectral responses of the pixels [Chakrabarti et al. 2014].

Previous demosaicking algorithms use this constraint in a nonlinear optimization framework and jointly solve the R, G and B for the entire image [Chakrabarti et al. 2014; Condat 2009]. However, it is infeasible to solve burst photography with color quanta images as an optimization problem since several thousands of binary images are involved. Instead, we first use the proposed joint demosaicking and merging algorithm to get four channels separately, and then use the linear constraint in Eq. 7 to scale the R, G and B channels at each pixel such that their sum is equal to W. This scale factor can be easily computed as the ratio between the measured W value and the expected W value from the measured R, G and B value:

$$k(x,y) = \frac{W(x,y)}{w_R R(x,y) + w_G G(x,y) + w_B B(x,y)}\,. \tag{8}$$

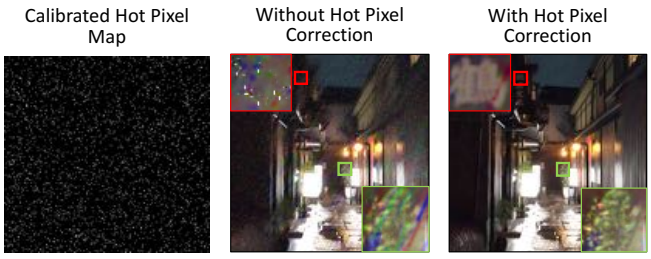

Fig. 5. **Correcting hot pixels. (Left)** 3% of total pixels are classified as hot pixels in our hardware prototype (DCR ≥30cps). **(Center)** Without hot pixel correction, the motion estimate is biased towards zero in dark regions, which results in blur. In bright regions, the hot pixels "move" together with the patches, which leaves color streaks in the result. **(Right)** With the proposed hot pixel correction, most of the blur and color streaks can be removed.

This two-step approach reduces the computational complexity significantly. As shown in Fig. 4, the merged result has sharper edges and less noise than the reconstruction from RGB channels only.

4.3 Handling Practical Challenges

Challenge of Hot Pixels in Quanta Images. Current SPAD arrays suffer from spatially-varying dark count rate (DCR), and especially, "hot pixels" which have exceptionally high DCR as compared to other pixels [Antolovic et al. 2016]. Identification and correction of hot pixels is especially important for low-light imaging since a DCR higher than or comparable to the actual light signal can significant downgrade the image quality. Fig. 5 (a) shows the calibrated hot pixel map of a SwissSPAD2 [Ulku et al. 2019], where about 3% of total pixels are identified as hot pixels. It is critical to remove hot pixels prior to motion estimation and alignment because hot pixels, if not alleviated, could strongly bias the motion estimate towards zero as they have extremely high intensities and do not move despite the presence of scene or camera motion. This is demonstrated in Fig. 5. Without hot pixel correction, the motion estimate is zero in the dark regions, causing motion blur. In the dark regions, the

Fig. 6. **Denoising with chrominance-focused BM3D.** By applying different level of denoising to the luminance and chrominance channels, chrominance-focused BM3D is able to suppress the color noise while maintain more structural details in the luminance channel.

hot pixels are warped together with the estimated motion, which causes color streaks in the reconstruction. The proposed hot pixel correction algorithm removes most of the blur and color streaks. **Please see the supplementary report for algorithm details.**

Chrominance-Focused Denoising. After temporal merging, we perform denoising to further improve the SNR by utilizing the spatial correlations. However, for single-photon cameras with pseudorandom CFAs, it is particularly challenging to find spatial correlations in individual binary frames. Therefore we apply merging and demosaicking first which generates an RGB image that is amenable to existing denoising techniques such as BM3D [Dabov et al. 2007].

Classical BM3D assumes the noise variance in the R, G and B channel is equal. The key difference in our setting is that, for RGBW color filter arrays, the luminance of an image has a much higher SNR than the chrominance. To leverage this benefit, we propose a *chrominance-focused denoising* approach: Prior to denoising, we convert the image into a modified YCbCr space [ITU-R 2011], which allows us to choose a smaller σ for the Y channel and apply more aggressive denoising to the Cb/Cr channels. Fig. 6 shows that the proposed chrominance-focused BM3D is able to suppress the color noise while maintaining structural details in the luminance channel. **More details can be found in the supplementary report.**

5 ACQUISITION OF COLOR: DESIGN AND ANALYSIS OF CFAS FOR SINGLE-PHOTON BURST PHOTOGRAPHY

While the previous section proposed computational and algorithmic approaches for processing the raw single-photon color frames, in this section, we consider the problem of raw frame acquisition itself. In particular, we address the following question: What is the right color filter array (CFA) [Adams et al. 1998] for color single-photon imaging? Our design of CFA is based on two main observations: First, the fine-grained motion measurement due to high frame-rate of SPADs could effectively increase the sampling rate of color, which allows using RGBW CFA patterns with large fraction of white pixels (thus increasing the overall SNR). Second, to further reduce color aliasing artifacts, we design a pseudorandom RGBW pattern which is inspired by previous blue-noise pattern designed for RGB pixels.

5.1 Burst Photography with RGBW Patterns

In addition to RGB patterns which are commonly used in commercial cameras, various RGBW patterns have been proposed [Chakrabarti 2016; Chakrabarti et al. 2014; Oh et al. 2017; Parmar and Wandell 2009]. The key idea is to add a fourth white pixel which is not color-selective and absorbs more light. Such higher light sensitivity improves the SNR of reconstructed images, especially relevant for low-light imaging. The density of RGB pixels is lowered in such RGBW patterns, which makes it necessary to design demosaicking algorithms accordingly [Bai and Li 2019; Oh et al. 2017].

Supersampling due to motion. One important observation for burst photography is that, when the inter-frame motion is correctly measured and compensated, the color samples measured by each color filter are dispersed along the motion trajectories. Therefore, the color information could be sampled at a higher spatial frequency than in the original CFA. This is especially pertinent for SPADs since pixels are sampled in time much more frequently, which corresponds to effectively higher color sampling rate.

To verify this observation, we simulate a challenging binary sequence with high-frequency content (fences). We use an existing RGBW pattern [Kwan et al. 2020] and apply the proposed algorithms (Section 4) to reconstruct an image, as shown in Fig. 7 (Left). A large amount of color artifacts can be observed when the scene is static (Top left). When the camera is moving horizontally (Middle left), the horizontal color sampling rate increases and the artifacts at the vertical fences are decreased (red crop). However, artifacts at edges with other orientations remain (yellow crop). When the camera is subject to random 2D motion (e.g., handheld motion) (Bottom left), the color sampling rate increases in both dimensions, and therefore artifacts at edges with all orientations disappear. To summarize, in the presence of motion, the proposed computational techniques are able to leverage the strength of RGBW filters to achieve high SNR in low light, while not getting adversely affected by the low color sampling rate. In practice, using our algorithmic pipeline, a CFA with 75% of W pixels achieves a good balance between light sensitivity and color sampling rate. **Please see the supplementary technical report for more details on CFA design and analysis.**

5.2 Periodic vs. Pseudorandom Patterns

Periodic CFAs such as Bayer patterns consist of blocks of fixed permutation of color filters (*e.g.*, RGGB) that are repeated spatially, while pseudorandom CFAs contain pseudorandomly generated pixels without repetitions [Chakrabarti 2016; Chakrabarti et al. 2014; Condat 2010; Oh et al. 2017; Sharif and Jung 2019]. The benefit of using pseudorandom patterns is that aliasing artifacts appear as incoherent noise, which is less perceptible than coherent moiré patterns generated by regular patterns. However, a completely random pattern makes demosaicking more challenging, leading to severe color artifacts. Inspired by existing work on pseudorandom RGB pattern [Condat 2010], we design a novel blue-noise RGBW pattern, as shown in Fig. 8 (f).

The performance of the proposed pattern can be analyzed in the frequency domain [Alleysson et al. 2005; Condat 2010]. This is shown in Fig. 9, where the CFA is decomposed into luminance, red-green (R-G) chrominance and blue-yellow (B-Y) chrominance and transformed to Fourier domain (For simplicity, we only show the R-G chrominance spectrum). When an image is captured with a

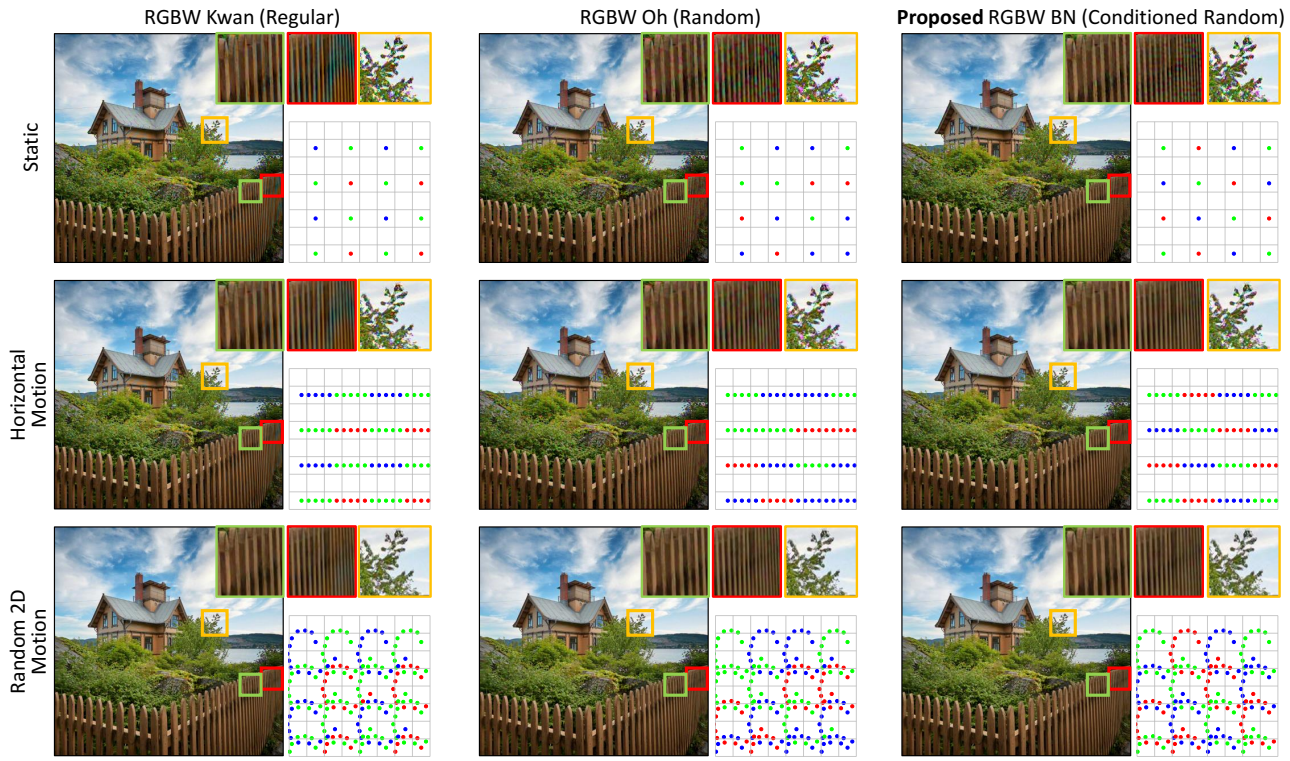

Fig. 7. **Color sampling with RGBW CFAs. (Left)** Simulated results with a regular RGBW pattern. **(Center)** Random RGBW pattern. **(Right)** Proposed conditioned random RGBW pattern. **(Top)** When the scene is static, the regular pattern introduces unnatural coherent moiré color artifacts. The random pattern generates less annoying incoherent artifacts, but becomes noticeable at higher frequencies (red crop). The conditioned random pattern generates fewer artifacts even at high frequencies. **(Middle)** When there is horizontal motion, artifacts around vertical edges (fences) are decreased, while artifacts around edges with other directions still remain. **(Bottom)** When random 2D motion is present, artifacts at edges with all directions are mitigated.

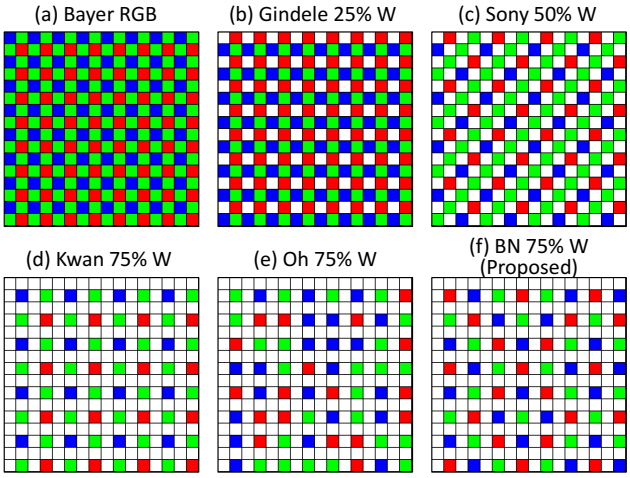

Fig. 8. **Related color filter arrays.** (a) Bayer RGB [Bayer 1976]. (b) 25% W [Gindele and Gallagher 2002]. (c) 50% W [Tachi 2012]. (d) 75% W regular [Kwan et al. 2020]. (e) 75% W random [Oh et al. 2017]. (f) 75% W blue-noise conditioned random (proposed).

CFA, the light signal arrives at the pixels (image irradiance) is convolved with the CFA, which, in the Fourier domain, corresponds to the multiplication of the irradiance spectrum with the CFA spectrum shown in the figure. For the Bayer pattern (Fig. 9(a)), the luminance energy is concentrated at $[0, 0]$, while the R-G chrominance energy is concentrated at high frequencies and therefore can be separated from the luminance during demosaicking for band-limited image radiance. A completely random pattern (*i.e.*, pixels generated from independent uniform distributions) distributes the chrominance energy uniformly (Fig. 9(b)), which makes the separation of luminance and chrominance difficult. [Condat 2010] proposed a blue-noise pseudorandom pattern which has an additional constraint that pixels of the same color cannot be contiguous. As a result, the chrominance has minimal energy in the baseband and therefore it can be separated from the luminance while avoiding structured, coherent aliasing artifacts, as shown in (Fig. 9(c)). We extend the blue-noise pattern to get an RGBW pattern with 75% W, as shown in Fig. 9(d). This periodic distribution of W pixels makes copies of the blur-noise spectrum and keeps the luminance and chrominance spectra separable. This regular arrangement of W pixels also makes inpainting the W channel easier and therefore eases alignment(Sec. 4.1).

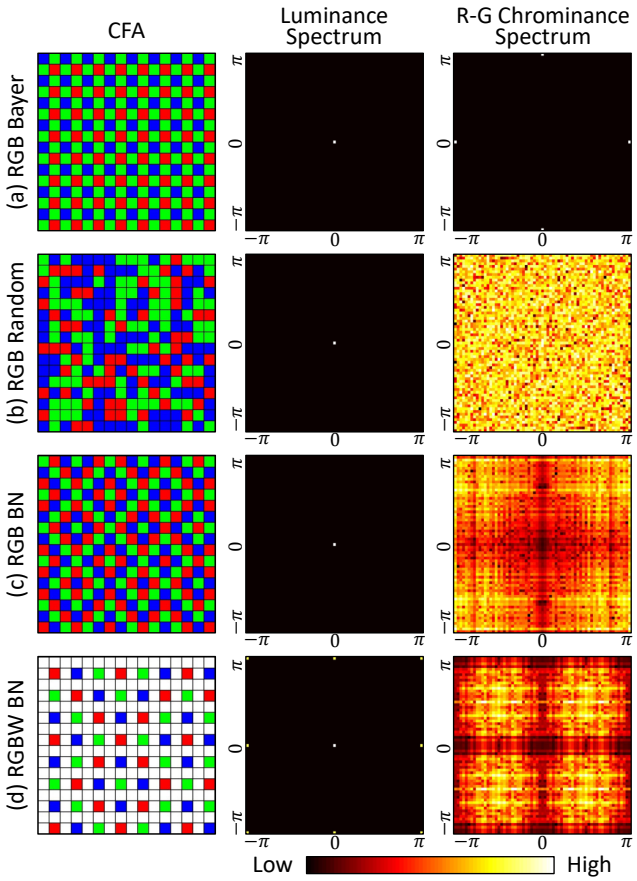

Fig. 9. **Spectrum analysis of periodic and pseudorandom patterns. (a)** RGB Bayer pattern: the luminance energy is concentrated at the origin, while the chrominance energy is concentrated at high frequencies, which means they can be separated well. **(b)** A completely random pattern has its chrominance energy uniformly distributed across all frequences, which makes it difficult to separate from luminance. **(c)** RGB blue-noise pattern makes luminance and chrominance separable while avoiding coherent aliasing artifacts. **(d)** We propose RGBW blue-noise by padding W pixels between color pixels, which retain the merits of the RGB blue-noise pattern.

We compare the proposed pattern with two existing 75% W CFAs with regular [Kwan et al. 2020] and completely random RGB filters [Oh et al. 2017] in Fig. 7. The blue-noise pattern improves the image quality for all three motion conditions, which is most prominent when the scene is static. The regular pattern introduces unnatural cyan/yellow moiré patterns in the image. The random pattern alleviates moiré patterns, but generates high-frequency color artifacts [Dippe and Wold 1985] (red crop). The proposed blue-noise RGBW pattern (conditioned random) ensures that aliasing between luminance and chrominance components is minimized even for high-frequency image content, as shown in (Top Right, red crop).

How much motion is needed? While the blue-noise pattern mitigate moiré patterns, such artifacts still exist for a completely static scene with high-frequency content. How much motion is required by the burst photography to remove them? Fig. 10(Left) plots the PSNR of the reconstructed image as a function of motion in pixels. 10 different sequences are synthesized using 2D circular rotation with different radii. It is clear that the biggest jump in PSNR comes when the motion is increased from 0 to 1 pixel. This is also demonstrated by Fig. 10(Center) and (Right): Most visible artifacts disappear when there is a 1-pixel motion. We conclude that the motion requirement is minimal for the proposed burst approach.

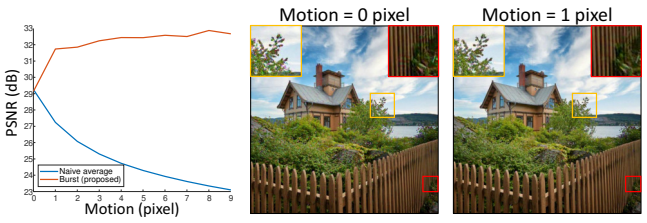

Fig. 10. **How much motion is needed? (Left)** We synthesize 10 different sequences using 2D circular rotation with different radii. The PSNR of naive averaging keeps decreasing due to motion blur. The PSNR of the proposed method keeps increasing, but the biggest jump comes when the motion is increased from 0 to 1 pixel. **(Center, Right)** Most visible artifacts are reduced or removed even when there is a 1-pixel motion.

Table 1. Sensor Parameters for Simulation

Sensor Type	Conventional	SPAD	Jot
Resolution	1024×1000	1024×1000	5120×5120
Frame Rate	240fps	24,000fps	1,000fps
Bit Depth	10	1	1
QE / PDE (R)	59%	9%	71%
QE / PDE (G)	65%	14%	79%
QE / PDE (B)	48%	13%	69%
Read Noise*	$2.4e^-$	0	$0.24e^-$
Dark Current Noise / Dark Count Rate*	$1e^-$/s	2.0cps	$0.16e^-$/s

**Note:* per-pixel.

6 RESULTS

6.1 Simulated Results

To evaluate the performance of the proposed approach under varying imaging scenarios in a controllable manner, we simulate the imaging process of single-photon sensors by using an open-source path tracer (Blender Cycles) to synthesize photorealistic scenes from high-quality 3D scene models, and then render mosaicked binary frames according to Eq. 3.

Conventional vs. SPAD color imaging under different lighting conditions. Fig. 11 shows a comparison between conventional CMOS-based burst photography and the proposed SPAD color imaging. Imaging parameters of the simulated sensors are summarized in Tab. 1. Parameters for the conventional CMOS sensor are adapted from a high-end machine-vision camera[2]. Parameters for the SPADs

[2]https://www.flir.com/products/grasshopper3-usb3/?model=GS3-U3-123S6C-C

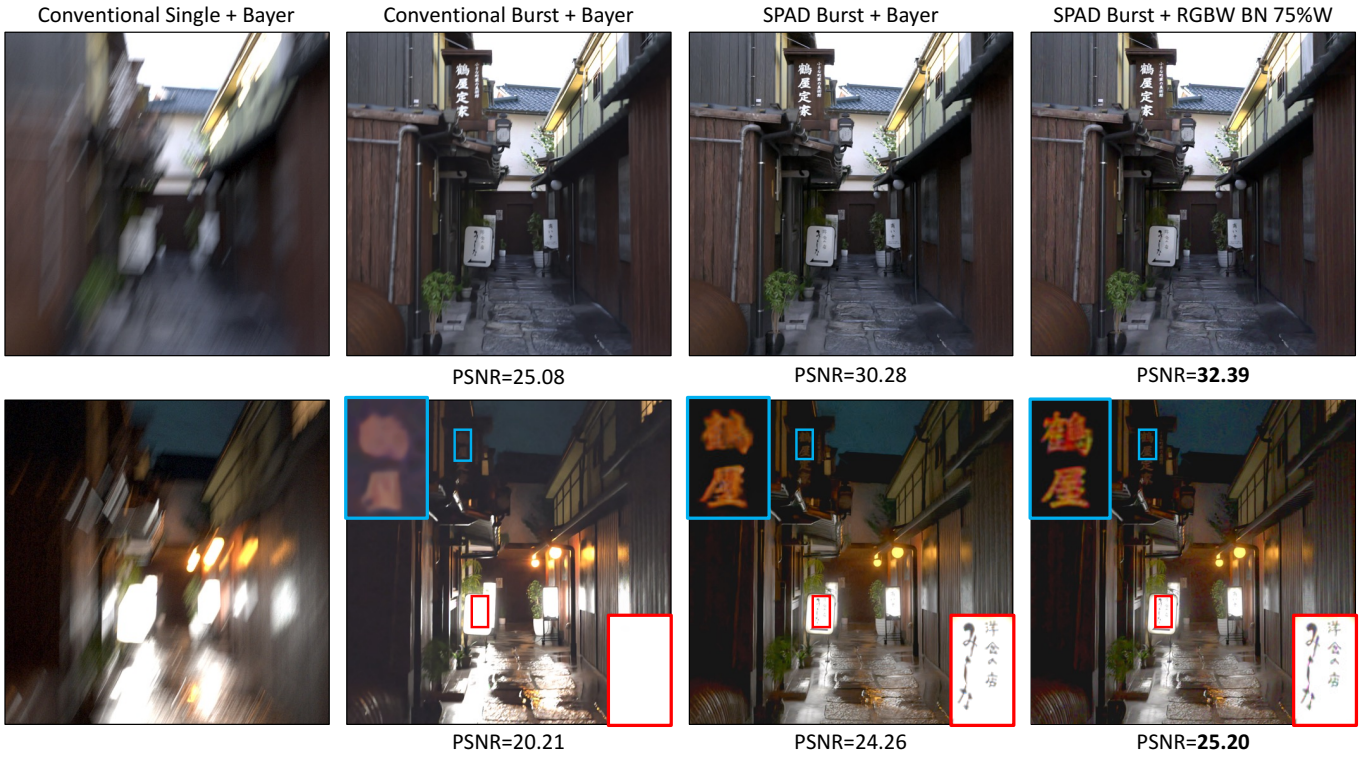

Fig. 11. **Comparison of conventional and SPAD burst photography under different lighting conditions on a simulated scene. (Top)** For a daytime scene with sufficient lighting, both conventional camera and SPAD are able to resolve the camera motion and reconstruct a blur-free, high-SNR image. **(Bottom)** For a night scene with both unlit regions and strong direct light sources, conventional burst photography cannot recover the entire dynamic range, while the proposed method can reconstruct both dark and bright regions. RGBW pattern recovers fine structure in the dark better than RGB pattern because of the higher transmission of W pixels. The same tone-mapping operator [Mantiuk et al. 2006] is applied to all results.

are based on [Morimoto et al. 2020], using a conservative estimate of effective fill factor (50%) which can be achieved with microlenses [Bruschini et al. 2023]. We adapt the proposed pipeline to process conventional images by skipping the motion interpolation within blocks. We keep the total exposure time same for both sensors, and compare them both visually and quantitatively by computing the PSNR of the linear reconstructions.

For a well-lit scene (Fig. 11 (Top)), both conventional camera and SPAD are able to resolve the camera motion and reconstruct a blur-free, high-SNR image. The same scene captured during nighttime (Fig. 11 (Bottom)) presents challenges, with both dark (unlit regions on the left) and bright parts (illuminated signs and lamps) creating a high dynamic range. Conventional burst photography cannot recover the entire dynamic range, resulting in noise and blur in the dark regions (blue window) and saturation in the bright regions (red window). On the other hand, the proposed method recovers both the dark and bright intensity ranges. Notice that the proposed RGBW pattern can better reconstruct the fine structure in the dark (text in the blue window) than conventional Bayer pattern due to the higher transmission of the white pixels.

Trade-off between spatial and temporal resolution. Fig. 12 compares the results of the proposed method on SPADs and jots, with the same camera motion and total exposure time. Simulated camera parameters are listed in Tab. 1. Recent development of jots-based quanta image sensors (CIS-QIS) has focused on spatial resolution instead of fast single-bit readout [Ma et al. 2022]. Here we assume a high-speed jot device optimized for fast burst photography and provide a projected set of parameters: We assume a frame rate of 1000 fps as reported in [Ma et al. 2017], and then match the total bandwidth to SPADs. We wish to emphasize that the goal is not to directly compare the two technologies, but to understand the trade-off between spatial and temporal resolution.

Due to their lower temporal resolution, jots cannot resolve motion blur under fast and complex camera motion. However, when the camera motion is slow, jots are able to recover sharper image details. We expect the two kinds of single-photon sensors to complement

Table 2. Quantitative Evaluation on Interpolated Video Data

	PSNR (↑)	SSIM (↑)	LPIPS (↓)
Naive average (long)	23.87	0.5934	0.5173
Naive average (short)	20.62	0.4361	0.6153
VBM4D [Maggioni et al. 2012]	21.20	0.5342	0.4382
MFIR [Bhat et al. 2021]	23.63	0.6163	0.4230
BIPNet [Dudhane et al. 2022]	25.25	0.6481	0.3357
Proposed	**26.06**	**0.7879**	**0.2665**

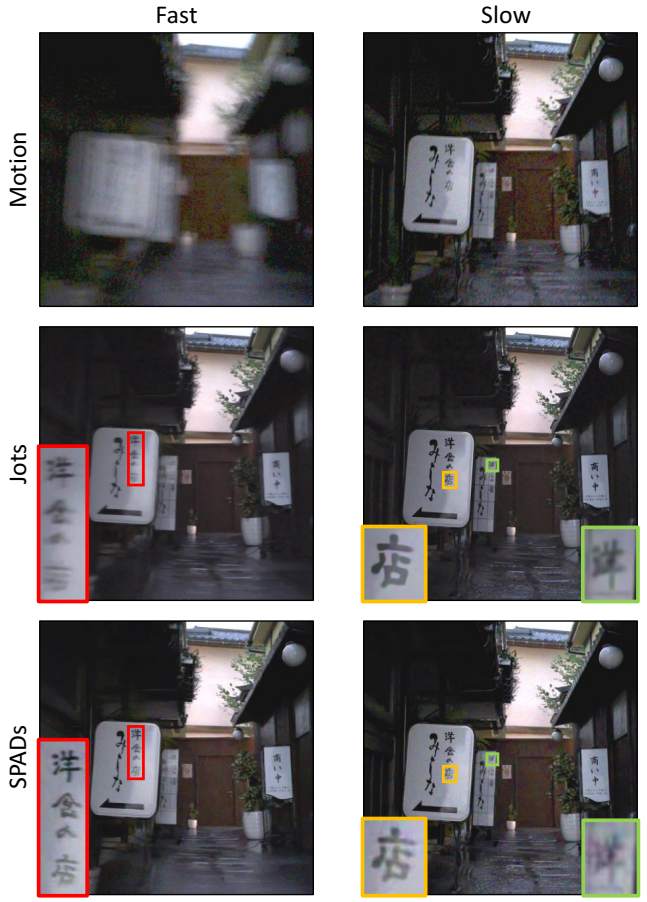

Fig. 12. **Comparison of simulated SPAD- and jot-based quanta burst photography.** A naive average of SPAD frames is shown to visualize the amount of motion. **(Left)** When the camera motion is extremely fast and complex, jots fail to resolve the motion blur. **(Right)** When the motion is slow, jots are able to recover more spatial image details.

each other in real applications, where SPADs are preferred when complex or high-speed motion is involved, and jots are preferred when high-frequency spatial details need to be recovered.

Quantitative evaluation on interpolated video data. In addition to comparison with other imagers, we also compare with existing burst denoising methods using synthetic SPAD data. To synthesize the high frame-rate data captured by a SPAD camera, we temporally interpolate a 1000FPS video dataset (X4K1000FPS [Sim et al. 2021], test set only) by a factor of 16x using RIFE [Huang et al. 2022] and then sample binary images according to Tab. 1. This gives us fifteen 512-frame binary sequences, with an average flux of 0.1 photons/pixel/frame, which is a very challenging scenario.

We compare with five baselines as shown in Fig. 13 and Tab. 2. Naive average simply takes the average of either the entire sequence (long) or a single block used by the proposed method (short). Burst photography algorithms that take raw images as input cannot be directly applied to our RGBW CFA. Instead, we first apply a universal demosaic algorithm [Condat 2009] on block-sum images and then run classic (VBM4D [Maggioni et al. 2012]) and pretrained learning-based (MFIR [Bhat et al. 2021] and BIPNet [Dudhane et al. 2022]) burst denoising algorithms. Naive average results in either significant motion blur (long) or noise (short). VBM4D smooths out the noise but leaves low-frequency noise patterns. MFIR and BIPNet cannot remove the noise perfectly without oversmoothing the structure (Fig. 13 only shows qualitative results for BIPNet as it performs better). In contrast, the proposed method is able to compensate for the motion while reconstructing a clean image, which achieves best PSRN, SSIM and LPIPS. Notice that the performance of learning-based methods can potentially be improved by optimizing on raw binomial images captured with the proposed CFA, which we leave for future work. **Details on this comparison can be found in the supplementary technical report.**

6.2 Hardware Prototyping a Color SPAD Array

We follow the CFA design principles as described in Sec. 5, and fabricate the 75% RGBW BN pattern on a SwissSPAD2 [Ulku et al. 2019] SPAD array as shown in Fig. 1. The pattern contains 75% W pixels and 8.33% R, G and B pixels each. With this hardware prototype, we are able to record 496×254 mosaicked binary images at up to 96.8kfps. To reduce the amount of data, we capture most of the scenes at 10kfps, and use 96.8kfps only for HDR scenes. Our prototype needs to be connected to two bench power supplies, limiting its portability. Therefore, we capture scenes that are representative of different imaging challenges and report qualitative results.

Color filters. The color filters are fabricated with photolithography, utilizing colored SU-8 photoresists [Jiang et al. 2020]. Fig. 14 shows the spectral response of the fabricated color filters. To evaluate the color reproduction performance, we use a Calibrite ColorChecker Classic with a D65 illuminant and plot the reference and measured color on chromaticity plots. Due to limitations in our fabrication process, in our current prototype, the color filters do not have ideal color selectivity, which results in images with low saturation. Furthermore, the blue filter has an unwanted strong response at wavelengths greater than 650nm, causing nonideal color reproduction performance. Fortunately, this is not a fundamental limitation of photolithography or SPADs, and can be solved with more iterations of photolithography experiments.

Performance under different light levels. Fig. 15 shows a scene under two lighting conditions, captured by a handheld SPAD camera moveing randomly in 3D. The proposed method reconstructs a low-noise, blur-free image, which again outperforms the four baselines. Note that some hot pixels are not completely removed by the preprocessing step and still remain in the baseline results where there is a cluster of them. The proposed robust merging step treats the remaining hot pixels as noise and can remove them. BIPNet results look blurrier since its input size is fixed to 8 frames, causing more intra-frame motion. The proposed method can interpolate the motion to each binary frame and resolve motion better.

Performance on challenging objects. Fig. 16 shows results on scenes with geometrically and radiometrically challenging objects, including high-frequency geometric structures (fence), complex occlusions (plant branches), specular reflection (fake fruits), and thin structures

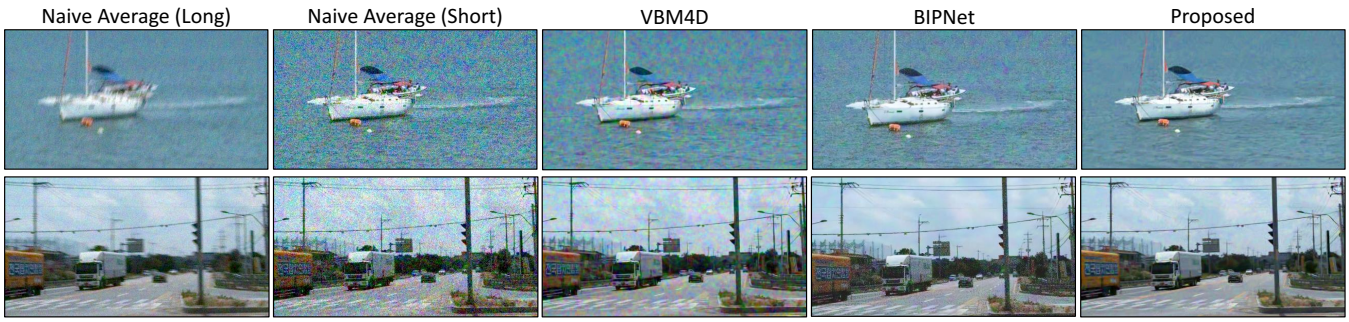

Fig. 13. **Qualitative comparison on interpolated video data.** Here we show two examples from the 15 interpolated sequences. Compared to baseline methods, the proposed method reconstructs the most clean and sharp image.

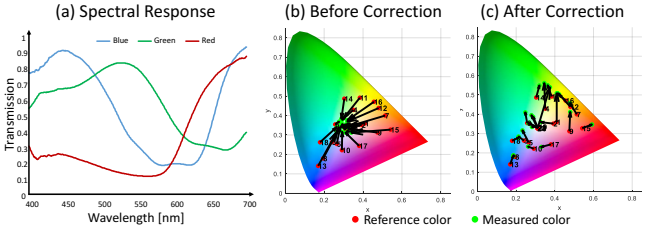

Fig. 14. **Color filters. (a)** Spectral response of the fabricated color filters. **(b)** We evaluate the color reproduction performance using a Calibrite ColorChecker Classic and a D65 illuminant. We plot the reference and measured color for each color patch in a chromaticity plot. **(c)** Chromaticity plot after applying an affine color correction matrix.

with depth variations. All these scenes pose stringent challenges for motion estimation and/or demosaicking algorithms. Nevertheless, the proposed method is able to reconstruct higher-quality images than the baseline methods.

High dynamic range. Fig. 17 shows two scenes with high dynamic range. The mosaicked binary frames show that the density of detected photons varies significantly across the image. To further illustrate the high dynamic range of the scene, we show the reconstructed linear images using the proposed method with different intensity scale factors. Fig. 17 (Top) demonstrates that the proposed method can reconstruct both the dark (artwork inside the room) and bright (sky and cloud) regions of the scene, which differ greatly in intensity such that they cannot be visualized simultaneously in a single 8-bit image. Fig. 17 (Bottom) shows an even more challenging example: We put several LED lights in a vase placed in a dark room. The proposed method can still reconstruct the entire dynamic range, including the pattern on the backdrop and the detailed reflection within the lights. The linear images need to be scaled by 500× to visualize the dark parts. For this scene, we also show the result of HDR+ [Hasinoff et al. 2016][3] on 20 DSLR images, which gives a sense of how commercial cameras perform for the same scene. DSLR recovers better colors, but lacks details in the bottom half. Notice this is not intended to be a direct comparison, as the DSLR has been heavily engineered over the years and has significantly higher resolution, quantum efficiency, *etc.*

[3]Open-source implementation: https://github.com/martin-marek/hdr-plus-swift

Complex scene motion. In addition to *rigid camera motion*, the color burst photography approach is also robust to fast, complex *scene motion*. Fig. 18 (Top) shows a rotating color wheel. We show mosaicked binary frames captured at two time instants to illustrate the motion. Unsurprisingly, simple averaging over a sequence blends the color due to motion blur. Despite the proposed method not explicitly modeling the rotation of objects, it can reconstruct a high-quality image with no color blending. Fig. 18 (Bottom) shows the non-rigid waving motion of a feather. The apparent motion varies considerably across pixels, which is difficult to estimate accurately, especially with the geometric complexity of thin structures. Nevertheless, the proposed method reconstructs a clean image with minimal motion blur. **Please see the supplementary material for the full videos, and more results.**

7 THEORETICAL ANALYSIS OF DYNAMIC RANGE

In Sec. 6, we demonstrate the HDR capability of the proposed color SPAD pipeline using qualitative experiment results. Can we quantify the dynamic range of color SPAD? While the dynamic range of monochromatic SPAD has been analyzed in the past [Antolovic et al. 2016; Ingle et al. 2021, 2019; Ma et al. 2020], there are important differences in the case of color SPADs. The distinctions are even more noteworthy for RGBW filter arrays.

First, consider the dynamic range of a single SPAD pixel. Assuming the scene motion is perfectly compensated, a closed-form expression for the SNR has been derived in previous work [Antolovic et al. 2016; Ingle et al. 2019; Ma et al. 2020]:

$$\mathrm{SNR} = \frac{\hat{\rho}}{\mathrm{RMSE}(\hat{\rho})} = \hat{\rho}\sqrt{\frac{n\tau}{e^{(\hat{\rho}+r_d)\tau}-1}}, \tag{9}$$

where $\hat{\rho}$ is the MLE of color intensity ρ (Eq. 6), n is the number of binary frames, τ is the exposure time for a single frame, and r_d is the dark count per frame. We define the dynamic range as the ratio between the maximum and the minimum measurable intensities:

$$\mathrm{DR} = 20\log_{10}\frac{\rho_{\max}}{\rho_{\min}}. \tag{10}$$

$\rho_{\max}$ and $\rho_{\min}$ are defined as the upper and lower bounds of ρ where the SNR is higher than 1. When the intensity is extremely low, the estimate is inaccurate due to the randomness of photon arrivals (shot noise). On the other hand, when the intensity is high, the SPAD is close to saturation and the intensity cannot be estimated reliably.

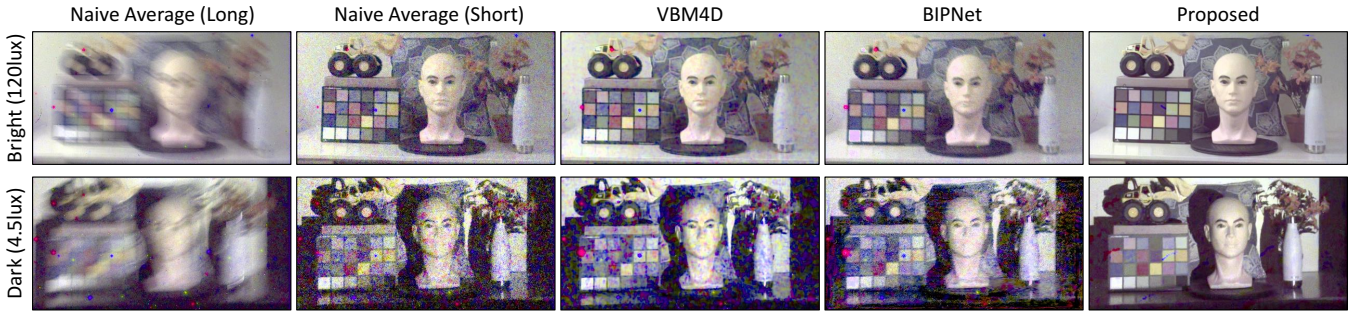

Fig. 15. **Performance under different light levels.** Baseline methods cannot fully remove the noise, especially in the dark scene (brightened 8.5X for visualization). In contrast, the proposed method reconstructs a clean, blur-free image. (Captured at 10kfps. 100 binary images per block, 20 blocks.)

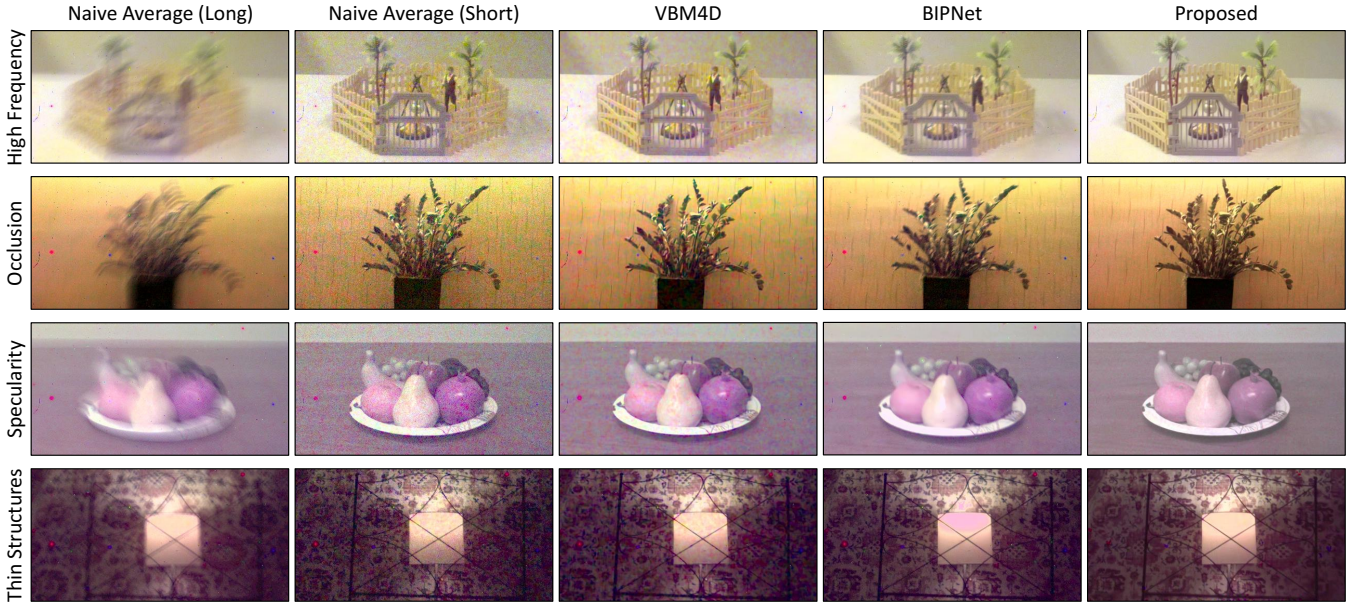

Fig. 16. **Performance on challenging objects.** Scenes with challenging objects, including high-frequency fence structures, complex occlusion between plant branches, specular reflection on fake fruits, and thin fence at a different depth than the rest of the scene. The proposed method outperforms naive average and the baseline methods. (Captured at 10kfps. 100 binary images per block, 20 blocks.)

The dynamic range can then be found by numerically solving the equation SNR = 1.

Next, consider our SPAD covered by RGBW color filters. Since pixels with different color filters have different PDE curves $\eta(\lambda)$ (Sec. 3), they have different SNRs even at the same incident photon flux level, resulting in different $\rho_{\max}$ and $\rho_{\min}$. Fig. 19 (Left) shows one example where the dynamic ranges of an R pixel and a W pixel are compared. The R pixel transmits less light and therefore works at a higher range of photon flux, while W pixel transmits more light and works for even less light but saturates more quickly. In other words, the dynamic ranges of pixels with different color filters are different. Based on this observation, we provide two different definitions of the dynamic range of the entire SPAD array:

Definition 7.1. **Best-performance DR** is defined as the dynamic range where all pixels (R, G, B and W) have SNR>= 1.

Definition 7.2. **Extended DR** is defined as the dynamic range where either the W pixels or the color pixels (R, G, B) have SNR>= 1.

Best-performance DR is given by the intersection of the dynamic ranges of different pixels, where all pixels give reliable estimate of the incident intensities. While best-performance DR gives a conservative definition of the dynamic range, it is worth noting that the photon flux range where only one type of pixel works still contains valuable information that can be used to generate an image.

- For the lower range where only the W pixel works, it is possible to reduce the saturation such that although the output image contains less color, it still carries reliable luminance information about the scene, which can be beneficial for both

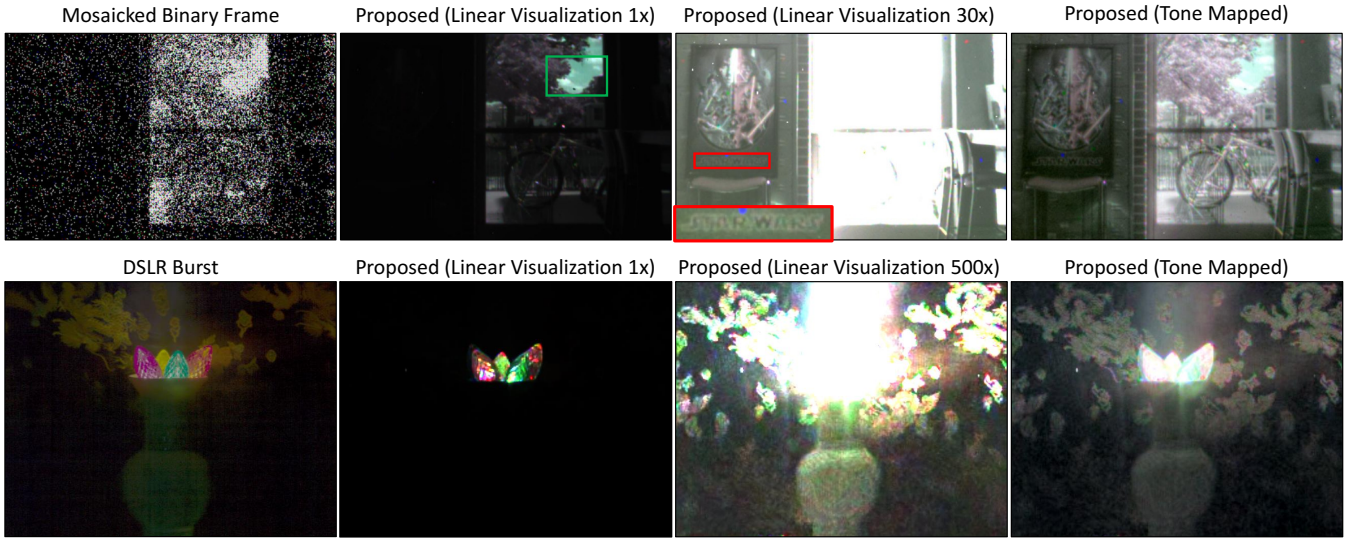

Fig. 17. **High dynamic range.** The proposed method is able to reconstruct scenes with high dynamic range. **(Top)** Both the sharp text in the dark and detailed shape of the cloud in the sky are reconstructed. **(Bottom)** Both the texture in the backdrop and the detailed reflection within the lights are reconstructed. 0.7 lux in the darkest region. (Captured at 96.8kfps. 2000 binary images per block, 30 blocks.)

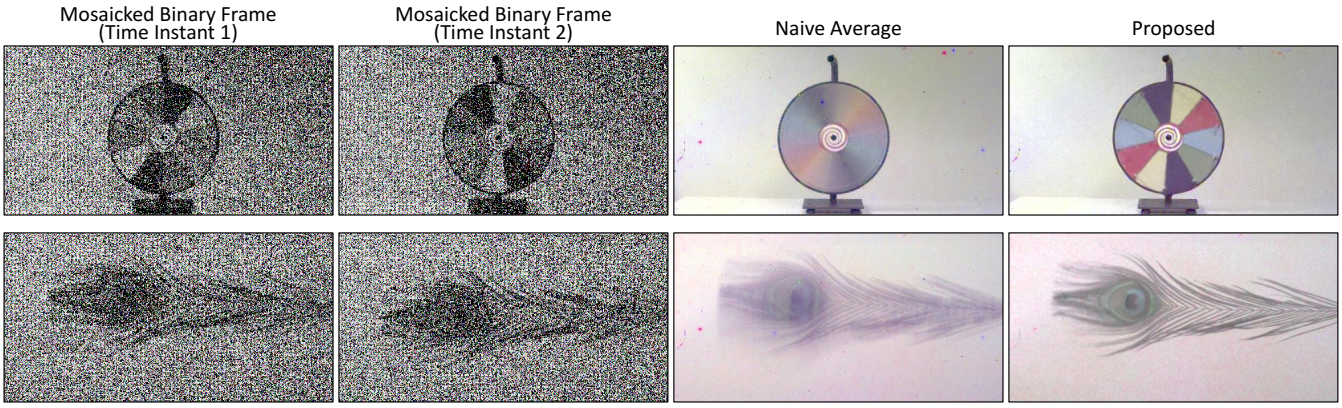

Fig. 18. **Complex scene motion.** In addition to rigid camera motion, the proposed method is also robust to complex scene motion. We show mosaicked binary frames captured at different time instants to visualize the scene motion. **(Top)** Naive average of a rotating color wheel blends the colors. The proposed method reconstructs a clear image with no color blending. **(Bottom)** Naive average of a waving feather creates motion blur. The proposed method is robust to this nonrigid, spatially-varying motion, and generates an image with significantly reduced motion blur. The blur on the feather tip cannot be perfectly removed due to faster motion. **Please refer to the supplementary video for a video reconstruction.**

recognition tasks in machine vision and artistic rendering for human viewing. Fig. 20 shows one example. Notice that the saturation is spatially-variant, which mimics human vision [Kirk and O'Brien 2011]. This approach can be considered a *noise-aware visualization technique*, **which we elaborate in the supplementary report.**

- For the higher range where only the RGB pixel works, it is possible to use the RGB pixel measurements only and do not include the contribution of W pixels during merging (Eq. 8) to avoid saturation.[4]

[4]Since alignment is still based on the W measurements, there is a possibility of introducing blur and ghosting artifacts at this higher range.

Putting all the three ranges together, we define the union of the dynamic ranges of different pixels as *extended DR*, which is the range where a reasonable image can be generated at the cost of less chrominance information in the darker range or possibility of blur and ghosting in the brighter range.

Trade-off between best-performance DR and extended DR.. The gap between best-performance DR and extended DR depends on the overlap between the DR of RGB pixels and W pixels, which in turn depends on the transmission of RGB filters. RGB filters with narrower passbands have more saturated colors and can reproduce the colors more faithfully, but the overall transmission is lowered. The transmission can be characterized by the coefficients w_R, w_G, w_B

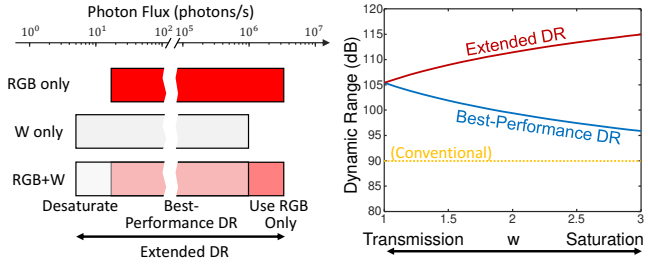

Fig. 19. **Dynamic range. (Left)** RGB pixels (Here we take R as an example) work at a higher flux range, while W pixels work at a lower flux range. We define the intersection of both ranges as the best-performance DR, and the union of both ranges as the extended DR. **(Right)** The trade-off between best-performance DR and extended DR is determined by whether transmission or saturation is preferred when choosing the color filters. When a color filter with higher saturation is chosen, the extended DR increases, while the best-performance DR decreases. Dynamic range of conventional burst photography with Bayer RGB is plotted for reference. Both conventional camera and SPAD camera are simulated using parameters in Tab. 1. An exposure time of 1s is used for both cameras.

Fig. 20. **Noise-aware visualization. (Left)** The original image contains significant color noise due to low light. **(Right)** The proposed noise-aware visualization reduces saturation of the dark pixels and suppresses noise.

as in Eq. 7; a higher value of w_R, w_G, w_B implies a lower transmission of the respective filter. We consider a simplified model where $w_R = w_G = w_B = w$ and plot the best-performance DR and extended DR in Fig. 19 (Right). For reference, we also plot the dynamic range of conventional burst photography with Bayer RGB based on the parameters in Tab. 1, which is independent of w. When the transmission decreases, the gap between RGB pixels and W pixels is widened, which means lower best-performance DR but higher extended DR. In practice, the transmission of the filters should be determined by considering whether best-performance DR or extended DR is preferred for a given end application.

8 LIMITATIONS AND FUTURE OUTLOOK

Computational complexity. Our unoptimized MATLAB implementation takes approximately 30 minutes to process 10,000 binary frames. The biggest bottleneck comes from the joint demosaic-merge step, which is not vectorized efficiently. From our experience with similar methods, it has been possible to achieve a speedup of 3-4 orders of magnitude by C++ implementation and optimization, which is an important next step towards a practical system.

Static scenes. The proposed method removes most of the color aliasing artifacts when there is at least a 1-pixel motion. Nevertheless, better demosaicking for completely static scenes can achieved by utilizing image priors learned by neural networks [Chakrabarti 2016; Sharif and Jung 2019]. Furthermore, the optimal RGBW CFA may also be learned as part of an end-to-end filter and network design for best image reconstruction quality [Chakrabarti 2016].

Optimal color filter design and implementation. Previous work has explored optimization of the spectral response of color filters for best color discrimination while reducing noise amplification [Kuniba and Berns 2009; Parmar and Reeves 2006]. Such optimization is an interesting future research direction. Specifically, an important future research question is: Should the color filters maximize transmission (for more light throughput) or maximize color saturation (better band selectivity)? This can be determined from the dynamic range analysis discussed in Sec. 7.

Learning-based burst photography. Learning-based burst photography algorithms for conventional CMOS cameras have achieved significant progress. However, learning-based methods are sensitive to out-of-distribution data and cannot be applied to single-photon images directly. A promising future direction is to develop neural networks-based methods for quanta burst photography by synthesizing quanta images from existing high-speed CMOS image datasets as well as capturing real datasets with SPAD cameras.

Comparison with commercial single-photon cameras. It is also worth mentioning that, compared to commercialized single-photon cameras[5], the proposed hardware is merely a research prototype that cannot compare directly in terms of quantum efficiency, resolution, color reproduction, *etc.* Nevertheless, the proposed techniques can be combined with the mature fabrication technologies that enable the commercial single-photon cameras, which can hopefully lead to high-performance color imaging systems that create quality images even for dynamic scenes with challenging dynamic ranges.

ACKNOWLEDGMENTS

This research was partially supported by NSF CAREER Award 1943149 and the Swiss National Science Foundation Grant 166289. The authors declare that there are no conflicts of interest related to this article. For the sake of transparency, the authors would like to disclose that (i) Edoardo Charbon holds the position of Chief Scientific Officer of Fastree3D, a company making LiDARs for the automotive market, and that (ii) Claudio Bruschini and Edoardo Charbon are co-founders of Pi Imaging Technology. Both companies have not been involved with the paper drafting, and at the time of writing have no commercial interests related to this article. Mohit Gupta is a co-founder of, and a stakeholder in Ubicept, Inc., which was not involved with the research performed for this publication.

[5]https://global.canon/en/news/2021/20211215.html, https://www.gigajot.tech/index.html

REFERENCES

J. Adams, K. Parulski, and K. Spaulding. 1998. Color Processing in Digital Cameras. *IEEE Micro* 18, 6 (1998), 20–30. https://doi.org/10.1109/40.743681

David Alleysson, Sabine Susstrunk, and Jeanny Herault. 2005. Linear Demosaicing Inspired by the Human Visual System. *IEEE Transactions on Image Processing* 14, 4 (April 2005), 439–449. https://doi.org/10.1109/TIP.2004.841200

Ivan Michel Antolovic, Samuel Burri, Claudio Bruschini, Ron Hoebe, and Edoardo Charbon. 2016. Nonuniformity Analysis of a 65-Kpixel CMOS SPAD Imager. *IEEE Transactions on Electron Devices* 63, 1 (Jan. 2016), 57–64. https://doi.org/10.1109/TED.2015.2458295

Chenyan Bai and Jia Li. 2019. Convolutional Sparse Coding for Demosaicking with Panchromatic Pixels. *Signal Processing: Image Communication* 77 (2019), 20–27.

Bryce E Bayer. 1976. Color imaging array. US Patent 3,971,065.

Goutam Bhat, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. 2021. Deep Reparametrization of Multi-Frame Super-Resolution and Denoising. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. 2460–2470. arXiv:2108.08286 [cs, eess]

Folkmar Bornemann and Tom März. 2007. Fast Image Inpainting Based on Coherence Transport. *Journal of Mathematical Imaging and Vision* 28, 3 (Oct. 2007), 259–278. https://doi.org/10.1007/s10851-007-0017-6

Claudio Bruschini, Ivan Michel Antolovic, Frédéric Zanella, Arin C Ulku, Scott Lindner, Alexander Kalyanov, Tommaso Milanese, Ermanno Bernasconi, Vladimir Pešić, and Edoardo Charbon. 2023. Challenges and Prospects for Multi-chip Microlens Imprints on Front-Side Illuminated SPAD Imagers. *Optica Open* (2023). https://doi.org/10.1364/opticaopen.22090487.v2

Claudio Bruschini, Harald Homulle, Ivan Michel Antolovic, Samuel Burri, and Edoardo Charbon. 2019. Single-Photon Avalanche Diode Imagers in Biophotonics: Review and Outlook. *Light: Science & Applications* 8, 1 (Dec. 2019), 87. https://doi.org/10.1038/s41377-019-0191-5

Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. 2015. Non-Line-of-Sight Imaging Using a Time-Gated Single Photon Avalanche Diode. *Optics Express* 23, 16 (Aug. 2015), 20997. https://doi.org/10.1364/OE.23.020997

Ayan Chakrabarti. 2016. Learning Sensor Multiplexing Design through Back-propagation. In *International Conference on Neural Information Processing Systems (NIPS'16)*. Curran Associates Inc., Barcelona, Spain, 3089–3097.

Ayan Chakrabarti, William T. Freeman, and Todd Zickler. 2014. Rethinking Color Cameras. In *IEEE International Conference on Computational Photography (ICCP)*. IEEE, Santa Clara, CA, USA, 1–8. https://doi.org/10.1109/ICCPHOT.2014.6831801

Stanley Chan, Omar Elgendy, and Xiran Wang. 2016. Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors. *Sensors* 16, 11 (Nov. 2016), 1961. https://doi.org/10.3390/s16111961

Paramanand Chandramouli, Samuel Burri, Claudio Bruschini, Edoardo Charbon, and Andreas Kolb. 2019. A Bit Too Much? High Speed Imaging from Sparse Photon Counts. In *IEEE International Conference on Computational Photography (ICCP)*. Tokyo, Japan, 1–9. arXiv:1811.02396

Chen Chen, Qifeng Chen, Minh N Do, and Vladlen Koltun. 2019. Seeing Motion in the Dark. In *International Conference on Computer Vision (ICCV)*. 3185–3194.

Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3291–3300.

Yiheng Chi, Abhiram Gnanasambandam, Vladlen Koltun, and Stanley H. Chan. 2020. Dynamic Low-Light Imaging with Quanta Image Sensors. In *European Conference on Computer Vision (ECCV)*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 122–138.

Joon Hee Choi, Omar A. Elgendy, and Stanley H. Chan. 2018. Image Reconstruction for Quanta Image Sensors Using Deep Neural Networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Calgary, AB, 6543–6547. https://doi.org/10.1109/ICASSP.2018.8461685

Laurent Condat. 2009. A Generic Variational Approach for Demosaicking from an Arbitrary Color Filter Array. In *IEEE International Conference on Image Processing (ICIP)*. IEEE, Cairo, Egypt, 1625–1628. https://doi.org/10.1109/ICIP.2009.5413388

Laurent Condat. 2010. Color Filter Array Design Using Random Patterns with Blue Noise Chromatic Spectra. *Image and Vision Computing* 28, 8 (Aug. 2010), 1196–1202. https://doi.org/10.1016/j.imavis.2009.12.004

Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. 2007. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Transactions on Image Processing* 16, 8 (Aug. 2007), 2080–2095. https://doi.org/10.1109/TIP.2007.901238

Mark A. Z. Dippe and Erling Henry Wold. 1985. Antialiasing Through Stochastic Sampling. *SIGGRAPH Comput. Graph.* 19, 3 (July 1985), 69–78. https://doi.org/10.1145/325165.325182

Xingbo Dong, Wanyan Xu, Zhihui Miao, Lan Ma, Chao Zhang, Jiewen Yang, Zhe Jin, Andrew Beng Jin Teoh, and Jiajun Shen. 2022. Abandoning the Bayer-filter to see in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17431–17440.

Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2022. Burst Image Restoration and Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5759–5768. arXiv:2110.03680 [cs]

Omar Elgendy, Abhiram Gnanasambandam, Stanley H. Chan, and Jiaju Ma. 2021. Low-Light Demosaicking and Denoising for Small Pixels Using Learned Frequency Selection. *IEEE Transactions on Computational Imaging* 7 (2021), 137–150. https://doi.org/10.1109/TCI.2021.3052694

Omar A. Elgendy and Stanley H. Chan. 2020. Color Filter Arrays for Quanta Image Sensors. *IEEE Transactions on Computational Imaging* 6 (2020), 652–665. https://doi.org/10.1109/TCI.2020.2964238 arXiv:1903.09823

Eric R. Fossum. 2005. What To Do With Sub-Diffraction Limit (SDL) Pixels?—A Proposal for a Gigapixel Digital Film Sensor (DFS). In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*. 214–217.

Eric R. Fossum. 2013. Modeling the Performance of Single-Bit and Multi-Bit Quanta Image Sensors. *IEEE Journal of the Electron Devices Society* 1, 9 (Sept. 2013), 166–174. https://doi.org/10.1109/JEDS.2013.2284054

Edward B Gindele and Andrew C Gallagher. 2002. Sparsely sampled image sensing device with color and luminance photosites. US Patent 6,476,865.

Abhiram Gnanasambandam and Stanley H. Chan. 2020. Image Classification in the Dark Using Quanta Image Sensors. In *European Conference on Computer Vision (ECCV)*. Springer, 484–501. arXiv:2006.02026

Abhiram Gnanasambandam, Omar Elgendy, Jiaju Ma, and Stanley H. Chan. 2019. Megapixel Photon-Counting Color Imaging Using Quanta Image Sensor. *Optics Express* 27, 12 (June 2019), 17298. https://doi.org/10.1364/OE.27.017298

Clément Godard, Kevin Matzen, and Matt Uyttendaele. 2018. Deep Burst Denoising. In *European Conference on Computer Vision (ECCV)*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Springer International Publishing, Cham, 538–554. https://doi.org/10.1007/978-3-030-01267-0_33

Alexander D Griffiths, Haochang Chen, David Day-Uei Li, Robert K Henderson, Johannes Herrnsdorf, Martin D Dawson, and Michael J Strain. 2019. Multispectral time-of-flight imaging using light-emitting diodes. *Optics express* 27, 24 (2019), 35485–35498.

Istvan Gyongy, Neale Dutton, and Robert Henderson. 2018. Single-Photon Tracking for High-Speed Vision. *Sensors* 18, 2 (Jan. 2018), 323. https://doi.org/10.3390/s18020323

Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. 2016. Burst Photography for High Dynamic Range and Low-Light Imaging on Mobile Cameras. *ACM Transactions on Graphics* 35, 6 (Nov. 2016), 1–12. https://doi.org/10.1145/2980179.2980254

Felix Heide, Steven Diamond, Matthias Nießner, Jonathan Ragan-Kelley, Wolfgang Heidrich, and Gordon Wetzstein. 2016. ProxImaL: Efficient Image Optimization Using Proximal Algorithms. *ACM Transactions on Graphics* 35, 4 (July 2016), 1–15. https://doi.org/10.1145/2897824.2925875

Felix Heide, Karen Egiazarian, Jan Kautz, Kari Pulli, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pająk, Dikpal Reddy, Orazio Gallo, Jing Liu, and Wolfgang Heidrich. 2014. FlexISP: A Flexible Camera Image Processing Framework. *ACM Transactions on Graphics* 33, 6 (Nov. 2014), 1–13. https://doi.org/10.1145/2661229.2661260

Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. 2022. Real-Time Intermediate Flow Estimation for Video Frame Interpolation. In *European Conference on Computer Vision (ECCV)*. Springer, 624–642. arXiv:2011.06294 [cs]

Atul Ingle, Trevor Seets, Mauro Buttafava, Shantanu Gupta, Alberto Tosi, Mohit Gupta, and Andreas Velten. 2021. Passive Inter-Photon Imaging. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 8585–8595.

Atul Ingle, Andreas Velten, and Mohit Gupta. 2019. High Flux Passive Imaging with Single-Photon Sensors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 6760–6769.

ITU-R. 2011. Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios. (2011).

Kiyotaka Iwabuchi, Yusuke Kameda, and Takayuki Hamamoto. 2021. Image Quality Improvements Based on Motion-Based Deblurring for Single-Photon Imaging. *IEEE Access* 9 (2021), 30080–30094. https://doi.org/10.1109/ACCESS.2021.3059293

Kiyotaka Iwabuchi, Tomohiro Yamazaki, and Takayuki Hamamoto. 2019. Iterative Image Reconstruction for Quanta Image Sensor by Using Variance-based Motion Estimation. In *International Image Sensor Workshop (IISW)*.

Haiyang Jiang and Yinqiang Zheng. 2019. Learning to see moving objects in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7324–7333.

Linan Jiang, Kyung-Jo Kim, Francis M. Reininger, Sebastien Jiguet, and Stanley Pau. 2020. Microfabrication of a Color Filter Array Utilizing Colored SU-8 Photoresists. *Applied Optics* 59, 22 (Aug. 2020), G137. https://doi.org/10.1364/AO.391579

Ahmet Serdar Karadeniz, Erkut Erdem, and Aykut Erdem. 2021. Burst photography for learning to enhance extremely dark images. *IEEE Transactions on Image Processing* 30 (2021), 9372–9385.

Fahad Shahbaz Khan, Rao Muhammad Anwer, Joost Van de Weijer, Andrew D. Bagdanov, Maria Vanrell, and Antonio M. Lopez. 2012. Color Attributes for Object Detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

IEEE, Providence, RI, 3306–3313. https://doi.org/10.1109/CVPR.2012.6248068
Adam G Kirk and James F O'Brien. 2011. Perceptually Based Tone Mapping for Low-Light Conditions. *ACM Transactions on Graphics* 30, 4 (2011), 10.
Filippos Kokkinos and Stamatis Lefkimmiatis. 2019. Iterative residual cnns for burst photography applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5929–5938.
Hideyasu Kuniba and Roy S. Berns. 2009. Spectral Sensitivity Optimization of Color Image Sensors Considering Photon Shot Noise. *Journal of Electronic Imaging* 18, 2 (2009), 023002. https://doi.org/10.1117/1.3116562
Chiman Kwan, Jude Larkin, and Bulent Ayhan. 2020. Demosaicing of CFA 3.0 with Applications to Low Lighting Images. *Sensors* 20, 12 (June 2020), 3423. https://doi.org/10.3390/s20123423
Jia Li, Chenyan Bai, Zhouchen Lin, and Jian Yu. 2017. Automatic Design of High-Sensitivity Color Filter Arrays With Panchromatic Pixels. *IEEE Transactions on Image Processing* 26, 2 (Feb. 2017), 870–883. https://doi.org/10.1109/TIP.2016.2633869
Zhetong Liang, Shi Guo, Hong Gu, Huaqi Zhang, and Lei Zhang. 2020. A Decoupled Learning Scheme for Real-world Burst Denoising from Raw Images. In *European Conference on Computer Vision (ECCV)*. 17.
Orly Liba, Ryan Geiss, Samuel W. Hasinoff, Yael Pritch, Marc Levoy, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qiurui He, Jonathan T. Barron, and Dillon Sharlet. 2019. Handheld Mobile Photography in Very Low Light. *ACM Transactions on Graphics* 38, 6 (Nov. 2019), 1–16. https://doi.org/10.1145/3355089.3356508
Ziwei Liu, Lu Yuan, Xiaoou Tang, Matt Uyttendaele, and Jian Sun. 2014. Fast Burst Images Denoising. *ACM Transactions on Graphics* 33, 6 (Nov. 2014), 1–9. https://doi.org/10.1145/2661229.2661277
Jiaju Ma, Saleh Masoodian, Dakota A. Starkey, and Eric R. Fossum. 2017. Photon-Number-Resolving Megapixel Image Sensor at Room Temperature without Avalanche Gain. *Optica* 4, 12 (Dec. 2017), 1474. https://doi.org/10.1364/OPTICA.4.001474
Jiaju Ma, Dexue Zhang, Dakota Robledo, Leo Anzagira, and Saleh Masoodian. 2022. Ultra-High-Resolution Quanta Image Sensor with Reliable Photon-Number-Resolving and High Dynamic Range Capabilities. *Scientific Reports* 12, 1 (Aug. 2022), 13869. https://doi.org/10.1038/s41598-022-17952-z
Sizhuo Ma, Shantanu Gupta, Arin C. Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. 2020. Quanta Burst Photography. *ACM Transactions on Graphics* 39, 4 (July 2020), 1–16. https://doi.org/10.1145/3386569.3392470
Matteo Maggioni, Giacomo Boracchi, Alessandro Foi, and Karen Egiazarian. 2012. Video Denoising, Deblocking, and Enhancement Through Separable 4-D Nonlocal Spatiotemporal Transforms. *IEEE Transactions on Image Processing* 21, 9 (Sept. 2012), 3952–3966. https://doi.org/10.1109/TIP.2012.2199324
Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2006. A Perceptual Framework for Contrast Processing of High Dynamic Range Images. *ACM Transactions on Applied Perception* 3, 3 (July 2006), 286–308.
Ben Mildenhall, Jonathan T. Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. 2018. Burst Denoising with Kernel Prediction Networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Salt Lake City, UT, 2502–2510. https://doi.org/10.1109/CVPR.2018.00265
Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul Srinivasan, and Jonathan T. Barron. 2021. NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images. *arXiv:2111.13679 [cs, eess]* (Nov. 2021). arXiv:2111.13679 [cs, eess]
Kazuhiro Morimoto, Andrei Ardelean, Ming-Lo Wu, Arin Can Ulku, Ivan Michel Antolovic, Claudio Bruschini, and Edoardo Charbon. 2020. Megapixel Time-Gated SPAD Image Sensor for 2D and 3D Imaging Applications. *Optica* 7, 4 (April 2020), 346–354.
K Morimoto, J Iwata, M Shinohara, H Sekine, A Abdelghafar, H Tsuchiya, Y Kuroda, K Tojima, W Endo, Y Maehashi, Y Ota, T Sasago, S Maekawa, S Hikosaka, T Kanou, A Kato, T Tezuka, S Yoshizaki, T Ogawa, K Uehira, A Ehara, F Inui, Y Matsuno, K Sakurai, and T Ichikawa. 2021. 3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing. In *IEEE International Electron Devices Meeting (IEDM)*. 20.2.1–20.2.4. https://doi.org/10.1109/IEDM19574.2021.9720605
Jun Ogi, Takafumi Takatsuka, Kazuki Hizu, Yutaka Inaoka, Hongbo Zhu, Yasuhisa Tochigi, Yoshiaki Tashiro, Fumiaki Sano, Yusuke Murakawa, Makoto Nakamura, and Yusuke Oike. 2021. A 124-dB Dynamic-Range SPAD Photon-Counting Image Sensor Using Subframe Sampling and Extrapolating Photon Count. *IEEE Journal of Solid-State Circuits* 56, 11 (Nov. 2021), 3220–3227. https://doi.org/10.1109/JSSC.2021.3114620
Paul Oh, Sukho Lee, and Moon Kang. 2017. Colorization-Based RGB-White Color Interpolation Using Color Filter Array with Randomly Sampled Pattern. *Sensors* 17, 7 (June 2017), 1523. https://doi.org/10.3390/s17071523
Yasuharu Ota, Kazuhiro Morimoto, Tomoya Sasago, Mahito Shinohara, Yukihiro Kuroda, Wataru Endo, Yu Maehashi, Shintaro Maekawa, Hiroyuki Tsuchiya, Aymantarek Abdelahafar, Shingo Hikosaka, Masanao Motoyama, Kenzo Tojima, Kosei Uehira, Junji Iwata, Fumihiro Inui, Yasushi Matsuno, Katsuhito Sakurai, and Takeshi Ichikawa. 2022. A 0.37W 143dB-Dynamic-Range 1Mpixel Backside-Illuminated Charge-Focusing SPAD Image Sensor with Pixel-Wise Exposure Control and Adaptive Clocked Recharging. In *IEEE International Solid- State Circuits Conference (ISSCC)*. IEEE, San Francisco, CA, USA, 94–96. https://doi.org/10.1109/ISSCC42614.2022.9731644
Manu Parmar and Stanley J. Reeves. 2006. Selection of Optimal Spectral Sensitivity Functions for Color Filter Arrays. In *International Conference on Image Processing*. IEEE, Atlanta, GA, 1005–1008. https://doi.org/10.1109/ICIP.2006.312669
Manu Parmar and Brian A. Wandell. 2009. Interleaved Imaging: An Imaging System Design Inspired by Rod-Cone Vision. In *IS&T/SPIE Electronic Imaging*, Brian G. Rodricks and Sabine E. Süsstrunk (Eds.). San Jose, CA, 725008. https://doi.org/10.1117/12.806367
Naama Pearl, Tali Treibitz, and Simon Korman. 2022. Nan: Noise-aware nerfs for burst-denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12672–12681.
Ximing Ren, Yoann Altmann, Rachael Tobin, Aongus Mccarthy, Stephen Mclaughlin, and Gerald S Buller. 2018. Wavelength-time coding for multispectral 3D imaging using single-photon LiDAR. *Optics express* 26, 23 (2018), 30146–30161.
Trevor Seets, Atul Ingle, Martin Laurenzis, and Andreas Velten. 2021. Motion Adaptive Deblurring with Single-Photon Cameras. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 1945–1954. arXiv:2012.07931
Yash D. Shah, Peter W. R. Connolly, James P. Grant, Danni Hao, Claudio Accarino, Ximing Ren, Mitchell Kenney, Valerio Annese, Kirsty G. Rew, Zoë M. Greener, Yoann Altmann, Daniele Faccio, Gerald S. Buller, and David R. S. Cumming. 2020. Ultralow-Light-Level Color Image Reconstruction Using High-Efficiency Plasmonic Metasurface Mosaic Filters. *Optica* 7, 6 (June 2020), 632. https://doi.org/10.1364/OPTICA.389905
S. M. A. Sharif and Yong Ju Jung. 2019. Deep Color Reconstruction for a Sparse Color Sensor. *Optics Express* 27, 17 (Aug. 2019), 23661. https://doi.org/10.1364/OE.27.023661
Hyeonjun Sim, Jihyong Oh, and Munchurl Kim. 2021. XVFI: eXtreme Video Frame Interpolation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. 14489–14498.
Masayuki Tachi. 2012. Image processing device, image processing method, and program pertaining to image correction. US Patent 8,314,863.
Hiroyuki Takeda, Sina Farsiu, and Peyman Milanfar. 2007. Kernel Regression for Image Processing and Reconstruction. *IEEE Transactions on Image Processing* 16, 2 (Feb. 2007), 349–366. https://doi.org/10.1109/TIP.2006.888330
Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolovic, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. 2019. A 512 × 512 SPAD Image Sensor With Integrated Gating for Widefield FLIM. *IEEE Journal of Selected Topics in Quantum Electronics* 25, 1 (Jan. 2019), 1–12. https://doi.org/10.1109/JSTQE.2018.2867439
Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. 2019. Handheld Multi-Frame Super-Resolution. *ACM Transactions on Graphics* 38, 4 (July 2019), 1–18. https://doi.org/10.1145/3306346.3323024 arXiv:1905.03277
Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. 2019. Basis Prediction Networks for Effective Burst Denoising with Large Kernels. *arXiv:1912.04421 [cs]* (Dec. 2019). arXiv:1912.04421 [cs]
Feng Yang, Y. M. Lu, L. Sbaiz, and M. Vetterli. 2012. Bits From Photons: Oversampled Image Acquisition Using Binary Poisson Statistics. *IEEE Transactions on Image Processing* 21, 4 (April 2012), 1421–1436. https://doi.org/10.1109/TIP.2011.2179306
F. Zappa, S. Tisa, A. Tosi, and S. Cova. 2007. Principles and Features of Single-Photon Avalanche Diode Arrays. *Sensors and Actuators A: Physical* 140, 1 (Oct. 2007), 103–112. https://doi.org/10.1016/j.sna.2007.06.021
Jing Zhao, Ruiqin Xiong, Hangfan Liu, Jian Zhang, and Tiejun Huang. 2021. Spk2ImgNet: Learning to Reconstruct Dynamic Scene from Continuous Spike Stream. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Nashville, TN, USA, 11991–12000. https://doi.org/10.1109/CVPR46437.2021.01182

Seeing Photons in Color: Supplementary Technical Report

SIZHUO MA, University of Wisconsin-Madison, USA and Snap Inc., USA
VARUN SUNDAR, University of Wisconsin-Madison, USA
PAUL MOS, CLAUDIO BRUSCHINI, and EDOARDO CHARBON, EPFL, Switzerland
MOHIT GUPTA, University of Wisconsin-Madison, USA

1 ALGORITHM DETAILS

In this section, we discuss the algorithm details that we exclude from the main paper to avoid distraction.

1.1 Estimating Motion for General CFAs

In the main paper, we discuss the need to convert mosaicked block-sum images into grayscale images before alignment. This approach can be applied to a general class of CFAs that only contain R, G, B and possibly W pixels. Specifically, we consider the following three cases, which are summarized in Fig. 1,

(1) For periodic CFA patterns with small periods (*e.g.*, 2×2 tiles, Fig. 1(a,b)), we downsample the mosaicked block-sum image by averaging each tile of filters [Hasinoff et al. 2016].
(2) For periodic patterns with large periods (Fig. 1(c)), the downsampling approach results in very low resolution which reduces alignment precision. In the extreme case, a pseudo-random pattern can be viewed as a pattern whose period is the size of the entire image. For such patterns, we choose our conversion strategy based on the fraction of W pixels
   (a) If the pattern contains a small fraction of W pixels (Fig. 1(c)), we notice it is possible to apply existing universal demosaicking algorithms [Condat 2009] and then convert the resulting RGB image into a grayscale image.
   (b) If the pattern contains a large fraction of W pixels (≥75%, Fig. 1(d,e,f)), the densely sampled W channel carries sufficient information for alignment. Therefore, we directly interpolate the W pixels [Bornemann and März 2007] to get a full-resolution grayscale image.

1.2 Reference Image for Merging

We first warp the grayscale block-sum images generated in the alignment step using the estimated block-level motion. The images are warped patch-wise using linear interpolation. Then these motion-compensated block-sum images are merged together using the Wiener-filter based approach proposed in [Hasinoff et al. 2016] to generate a grayscale reference image that is robust to alignment errors. Notice that the warping is applied on the block level as full-resolution grayscale images are not available at frame-level, which means the generated reference image may still contain blur due to intra-block motion. Nevertheless this reference image provides sufficient structure for guiding the merging of color samples in an edge-preserving, misalignment-resilient way.

Authors' addresses: Sizhuo Ma, sizhuoma@cs.wisc.edu, University of Wisconsin-Madison, USA, Snap Inc., USA; Varun Sundar, vsundar4@wisc.edu, University of Wisconsin-Madison, USA; Paul Mos, paul.mos@epfl.ch; Claudio Bruschini, claudio.bruschini@epfl.ch; Edoardo Charbon, edoardo.charbon@epfl.ch, EPFL, Switzerland; Mohit Gupta, mohitg@cs.wisc.edu, University of Wisconsin-Madison, USA.

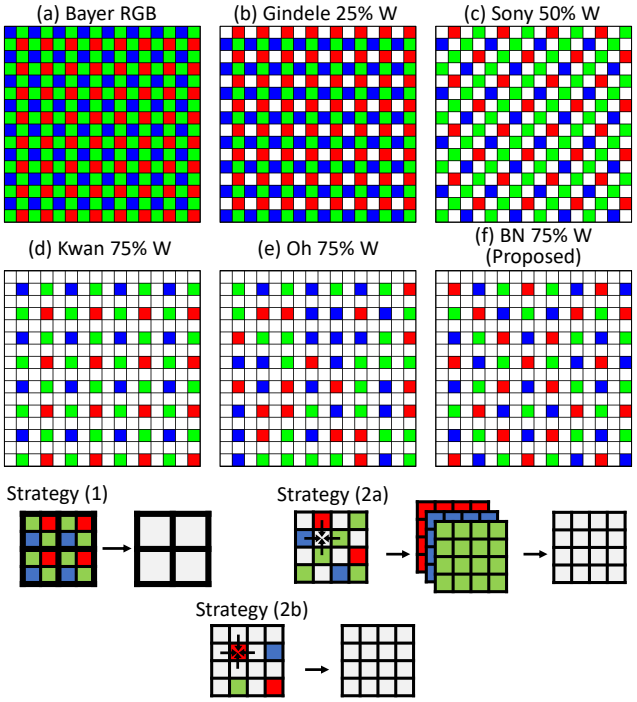

Fig. 1. **Estimating motion for general CFAs** We propose three different strategies for converting mosaicked block-sum images into grayscale images.

1.3 Robust Merging of Color Samples

As mentioned in the main paper, a joint demosaic-merge algorithm has to be used to combine the mosaicked quanta images to get an RGB image. This is done by treating each pixel in a mosaicked quanta image as a color sample, which belongs to one of the R, G, B or W channel. Color samples from different frames are warped to a common pixel grid (corresponding to the reference frame) using the estimated sub-pixel frame-level motion. The intensity value at each pixel is then reconstructed by taking a weighted sum of color samples within a neighborhood:

$$S\_(x,y) = \frac{\sum_{i \in \mathcal{N}} w_i \cdot S\_i}{\sum_{i \in \mathcal{N}} w_i}, \qquad \_ = \text{R, G, B or W}, \tag{1}$$

where $\mathcal{N}$ is the set of all sample points in the neighborhood around pixel (x, y). S_i is the value of the i-th sample point (0 or 1). w_i is the weight of the color sample which consists of two parts:

$$w_i = w_{Gi} \cdot w_{Ri}, \tag{2}$$

where w_{Gi} is given by anisotropic Gaussian kernel and w_{Ri} is given by a sample-wise robustness term to penalize misaligned patches.

Anisotropic Gaussian kernel. The anisotropic Gaussian kernel is given by the following equation,

$$w_{Gi} = \exp\left(-\frac{1}{2}(\mathbf{x_i} - \mathbf{x})^T \Omega^{-1} (\mathbf{x_i} - \mathbf{x})\right), \tag{3}$$

where $\mathbf{x} = (x, y)$ is the pixel location of interest, $\mathbf{x_i} = (x_i, y_i)$ is the location of the sample point after warping. The main purpose of using an anisotropic Gaussian kernel is to adaptively combine the color samples based on the local structure tensor of the reference image, which is encoded in the covariance matrix Ω. A larger kernel is used for denoising a flat region. A smaller kernel is used to preserve the high-frequency details of a textured region. An elongated kernel is used along edges to preserve the edge structure. We adopt the same kernel design as in [Ma et al. 2020].

Sample-wise robustness term. Each color sample is weighted by a robustness function so that color samples that are misaligned are assigned a lower weight to avoid artifacts. However, it is difficult to determine if a binary sample is misaligned or not. Therefore, we take an approach that is similar to the block-wise alignment step: We compute a weight function from the intensities of block-sum images, which is then broadcast to all the frames that constitute the block. Specifically, the weight is computed by comparing the grayscale block-sum images and the reference image,

$$R = \text{clamp}(s \cdot \exp\left(\frac{(x - \mu_s)^2}{s_c(\sigma_s^2 + \sigma_b^2)}\right), 0, 1), \tag{4}$$

where x is the pixel value in the block-sum image. s is a scale factor that depends on local motion variation M (highest magnitude difference of motion within a 3×3 neighborhood):

$$s = \begin{cases} 2 & \text{if } M > 10 \\ 12 & \text{otherwise} \end{cases} \tag{5}$$

If a pixel is at the discontinuity of the estimated motion field, it is more likely to be misaligned and large difference between block-sum images and the reference image is less permissible.

μ_s and σ_s are the mean and standard deviation of the intensities in a local 3×3 neighborhood in the reference image. If a pixel value is too different from this local intensity distribution in the reference frame, it is probably misaligned and will be assigned a lower weight during merging. Notice that since the block-sum image is the sum of a small number of binary images (typically 100), it suffers from the random noise of photon arrivals, which can be characterised by the binomial imaging model of single-photon cameras. Therefore, the weight function is corrected by another term σ_b, which is an estimate of standard deviation of the underlying binomial distribution:

$$\sigma_b = \frac{1}{T} \cdot \frac{S}{T} \cdot (1 - \frac{S}{T}), \tag{6}$$

where T is the number of frames in the block, and S is the pixel value (photon counts) in the reference frame. $s_c = 2$ is another scale factor that can be tuned to control the level of robustness. This weight function is computed from the W channel only and shared with the R, G and B channels because the W channel has a higher SNR.

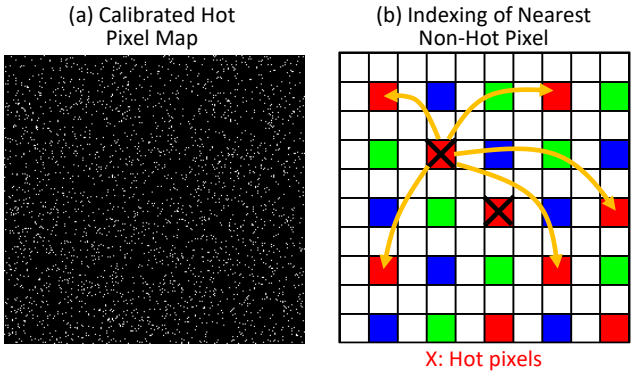

Fig. 2. **Correcting hot pixels. (a)** 3% of total pixels are classified as hot pixels in our hardware prototype (DCR threshold=30cps). **(b)** We propose a method for correcting hot pixels by replacing hot pixels in the mosaicked quanta images with the a random pixel in its k-nearest non-hot pixel neighbors. This list of k-nearest neighbors can be pre-computed for every hot pixel in the manufactured color SPAD array.

1.4 Correction of DCR Non-Uniformity

As mentioned in the main paper, correction of hot pixels must be performed on the raw mosaicked binary images themselves. Conventional image filters such as median filters cannot be applied to mosaicked binary images directly. We propose a *random replacement* approach: For each binary frame, we replace the binary value at the hot pixels by a random pixel in the neighborhood that has the same color filter. This effectively replaces the estimated intensity at the pixel by a spatial-temporal average of its neighbors. For efficiency, we pre-compute and store a list of nearby pixels with the same color for each hot pixel such that we only need to randomly pick a pixel from that list. This is illustrated in Fig. 2.

Previous image reconstruction methods for SPADs also discuss about how to correct the spatially non-uniform DCR. [Antolovic et al. 2016] calibrates and subtracts the DCR from each pixel in the estimated photon flux. In quanta burst photography, the non-uniformity of DCR is further complicated by motion: The spatial statistics of DCR are not preserved in the final image due to the motion compensation. In practice, we notice that the proposed robust merging algorithm is able to handle small variations of DCR and generate visually pleasing images. Therefore we do not explicitly correct for small variations of DCR but only remove hot pixels as a pre-processing step.

1.5 Chrominance-Focused Denoising

To leverage the benefit that W channels have higher SNR than RGB channels, we propose a *chrominance-focused denoising* approach: We first convert the image into a modified YCbCr space [ITU-R 2011] which is defined as:

$$\begin{aligned} Y &= k_R R + k_G G + k_B B, \\ Cb &= \frac{1}{2(1 - k_B)}(B - Y), \\ Cr &= \frac{1}{2(1 - k_R)}(R - Y), \end{aligned} \tag{7}$$

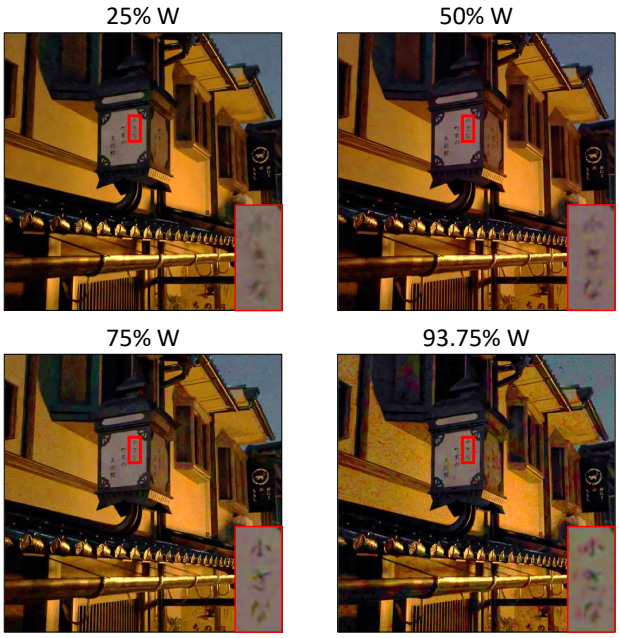

Fig. 3. **Simulated results for CFAs with different fractions of W pixels.** As the fraction of W pixels increases, the proposed method can reconstruct more spatial details in the image (clearer text in the red window), but also generate more color noise and artifacts. We find that 75% of W pixels achieves a good balance between light sensitivity and color sampling.

by choosing (k_R, k_G, k_B) as a normalized version of (w_R, w_G, w_B):

$$(k_R, k_G, k_B) = \frac{1}{\sqrt{w_R^2 + w_G^2 + w_B^2}} (w_R, w_G, w_B)\,. \tag{8}$$

Through this construction of YCbCr color space, Y channel is a scaled version of the reconstructed W channel such that the noise in the W pixels and RGB pixels are now separated in the luminance channel (Y) and the chrominance channels (Cb, Cr). In the grouping step of BM3D, patches are matched by their Y channel, and then the same grouping is applied to Cb and Cr channels. In the filtering step, the parameter σ (assumed noise standard deviation) is empirically chosen for the three channels. Since the luminance channel comes from W pixels and has a higher SNR than the chrominance channels, we choose a smaller σ for the luminance channel and apply more aggressive denoising to the chrominance channels.

We follow previous denoising practices for single-photon cameras [Chan et al. 2016; Gnanasambandam et al. 2019]: We first apply an Anscombe transform to convert the binomial-distributed multi-bit image into a Gaussian-distributed image with fixed variance [Anscombe 1948], and then apply chrominance-focused BM3D for denoising. After denoising, the inverse Anscombe transform is applied and Eq. 6 in the main paper is used to convert the photon counts to linear intensities.

2 CFA DESIGN DETAILS

2.1 What Is the Right Fraction of W Pixels?

RGBW CFAs with different proportions of W pixels have been proposed, from 25% [Gindele and Gallagher 2002] to over 90% [Sharif and Jung 2019]. Increasing the proportion of W pixels further improves the light sensitivity in the dark, but makes it harder to accurately recover the colors. Fig. 3 shows simulated result of the proposed pipeline in a low-light environment with different fraction of W pixels. As the fraction of W pixels increases, more spatial details in the image can be recovered (clearer text), but color noise and artifacts also increase. The optimal fraction of W pixels also depends on the exact algorithm being utilized. In practice, we find that, using our burst photography pipeline, 75% of W pixels achieve a good balance between light sensitivity and color sampling.

2.2 Could Spectral Multiplexing Help with SPAD Imaging?

Another thread of work has been focused on designing CFAs that multiplexes the three color channels in the frequency domain to achieve a balance between the light sensitivity, color aliasing and other factors [Bai et al. 2016; Elgendy and Chan 2020; Henz et al. 2018; Hirakawa and Wolfe 2008]. Each individual pixel is not limited to R, G, B or W but any convex combination of RGB. Optimal design can then be determined by solving an optimization problem. Since SPADs generate virtually no read noise, previous computational imaging theory has shown that multiplexing does not improve the SNR of RGB color estimates [Cossairt et al. 2013]. Nevertheless, increased light sensitivity can help alignment in the dark (in the same way as W pixels do), which we leave for future work.

3 ADDITIONAL EXPERIMENTAL RESULTS

Details on the quantitative evaluation on interpolated video data. We take the test set of X4K1000FPS [Sim et al. 2021], which contains fifteen 4K videos at 1000 FPS. Each video contains 32 frames, which we temporally interpolate by a factor of 16x using RIFE [Huang et al. 2022]. We spatially downsample the frames to 512x256 due to the memory constraints of BIPNet. We then a sample binary image from each intensity frame, which gives us fifteen 512-frame binary sequences.

We run the official v1.0 implementation of VBM4D[1]. We set sigma=-1, which lets the algorithm automatically picks the right sigma for denoising. We run the official implementation of MFIR[2] and BIPNet[3]. For the input noise variance map required by the networks, we estimate the variance of the linear flux estimator [Ma et al. 2020] which is then multiplied by a global scale factor. The global scale factor is chosen to be the largest value that does not oversmooth images or create artifacts.

Conventional vs. SPAD color imaging for an HDR scene. Fig. 4 gives another example of an HDR scene. A single exposure with a conventional sensor results in an underexposed image (1x exposure) or a blurred image with no details in the bright regions (500x exposure). By taking a sequence of 20 images, conventional burst photography

[1] http://www.cs.tut.fi/ foi/GCF-BM3D
[2] https://github.com/goutamgmb/deep-rep
[3] https://github.com/akshaydudhane16/BIPNet

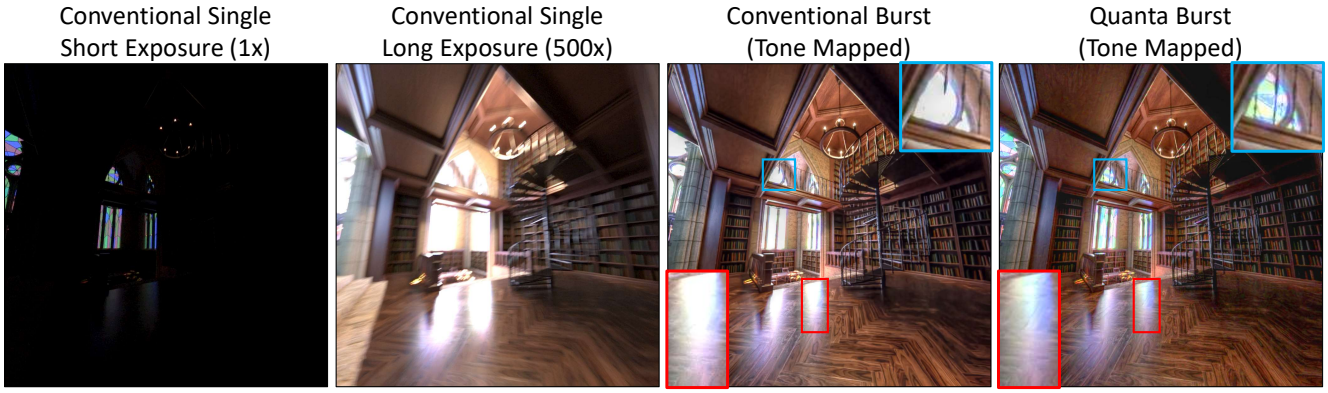

Fig. 4. **Performance for a simulated indoor HDR scene.** Conventional single-shot imaging results in either an underexposed image (1x exposure) or a blurred image with no details in the bright regions (500x exposure). Conventional burst photography extends the dynamic range, while the proposed method captures a even higher dynamic range and reconstructs both the texture on the floor and the patterns on the stained glass.

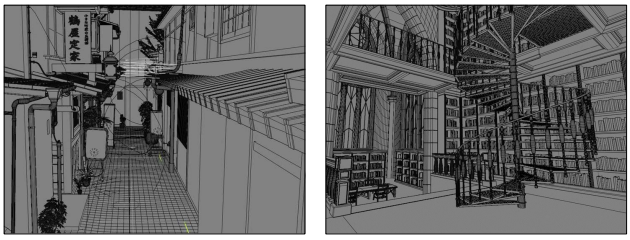

Fig. 5. **3D models used in the simulation.**

extends the dynamic range of the output. The proposed method, with the same total capture time, recovers a considerably higher dynamic range and reconstructs both the texture and the color gradients on the highlight on the floor, as well as the patterns on the stained glass.

Super-resolution. The proposed method is capable of super-resolution by choosing a pixel grid that has a higher resolution than the input image during the merge stage, which takes advantage of the sub-pixel motion between binary images. Fig. 6 shows an example. By reconstructing the image at 2X resolution, the image clearly reconstructs the digits on the dart board, which is illegible in naive average images and 1X reconstruction.

Nonrigid scene motion. Fig. 7 shows the reconstruction results of a person waving a cloth. Notice this motion is highly nonrigid, with complex wrinkle patterns on the cloth. Nevertheless, the proposed method is able to reconstruct a clean image of the cloth with minimal blur, as compared to the noisy results of the baseline methods. **Please see the supplementary video for a video reconstruction.**

Natural-looking scenes with HDR.. In addition to lab environments, Fig. 8 shows three natural-looking scenes with high dynamic range. By applying existing tone-mapping operators [Mantiuk et al. 2006], the proposed method can capture and reconstruct details in both bright regions and dark regions that are not direct illuminated, even if the camera is shaking.

4 VISUALIZING LOW-LIGHT COLOR IMAGES

So far we have discussed how to recover a linear measurement of irradiance on the single-photon image sensor. Like conventional digital photography with CMOS sensors, such linear images are not directly used for display but passed through a series of nonlinear process first such as tone mapping and gamma correction.

One especially challenging scenario for visualizing the captured color images is that, when the images are captured in extremely low light, they often contain severe, unpleasant looking noise that cannot be completely removed by denoising algorithms. Our key observation is that fortunately, with RGBW color filters, the amount of noise in the luminance channel is considerably lower than in the chrominance channel since W pixels have a higher light sensitivity. Based on this observation, we propose a visualization / display scheme tailored for color images captured in low-light using RGBW CFAs. The goal of the proposed method is to improve the visual quality of such low-light images by increasing the contribution of the luminance channel relative to the chrominance channel. Formally, we replace the color ratio of the R, G and B channels (dividing R, G and B by measured W channel) by a linear interpolation of the input image and a neutral gray color $\mathbf{p}$:

$$(\frac{R_{\text{vis}}}{W}, \frac{G_{\text{vis}}}{W}, \frac{B_{\text{vis}}}{W}) = \rho \cdot (\frac{R}{W}, \frac{G}{W}, \frac{B}{W}) + (1-\rho) \cdot \mathbf{p}\,. \tag{9}$$

ρ controls the interpolation weight: When $\rho = 1$, the output is the original image. When $\rho = 0$, no chrominance information is retained and a completely gray image is generated. $\mathbf{p}$ can be set to $(1/3, 1/3, 1/3)$. To simulate human vision, it is also possible to choose $\mathbf{p}$ by measuring how a gray patch on a color checker looks in low light, which shifts to a dull purple color [Jacobs et al. 2015]. Similarly inspired by previous work on simulating human vision [Jacobs et al. 2015], the weight ρ depends linearly on the log-luminance $\log W$:

$$\rho(x,y) = \text{clamp}(\frac{\log W(x,y) + L_r - C_2}{C_1 - C_2}, 0, 1)\,, \tag{10}$$

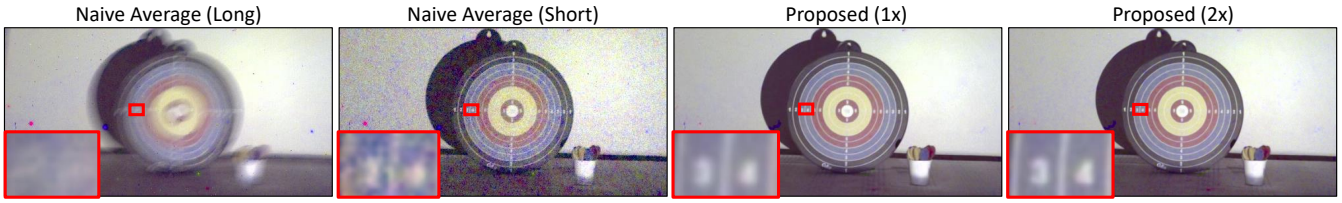

Fig. 6. **Super-resolution.** By reconstructing the image at 2X resolution, more image details are revealed and the digits on the dart board become readable. (Captured at 10kfps. 100 binary images per block, 20 blocks.)

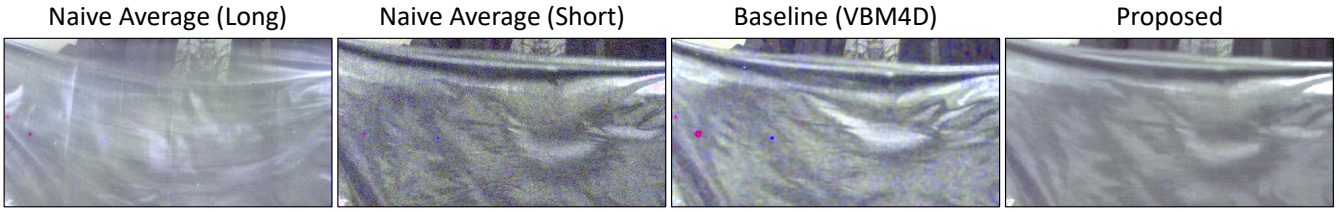

Fig. 7. **Nonrigid scene motion.** The proposed method works even when the highly nonrigid motion of the cloth is present, and reconstructs sharper and cleaner images than the baselines.

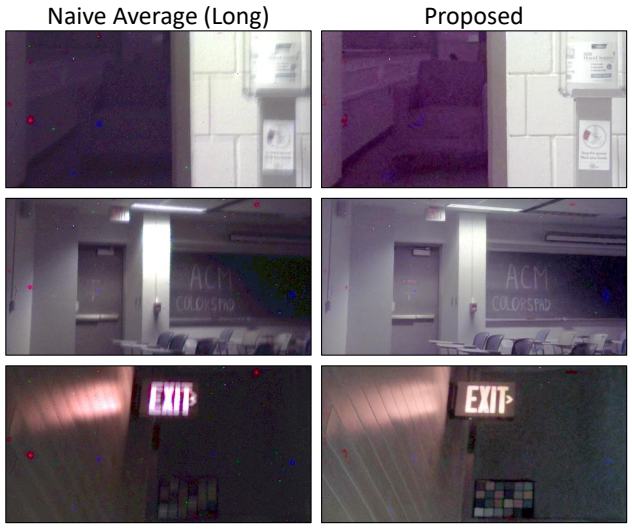

Fig. 8. **Natural-looking scenes with HDR.** The proposed method can capture and reconstruct image details in both bright regions and dark regions that are not directly under the lights, even with a moving camera.

where $C_1 = 0$, $C_2 = -2$ specifies the log-luminance range where this blending of color is in effect. $\rho(x, y)$ is computed at each pixel, which means the weight of chrominance varies across the image and is higher at brighter regions. We also provide the user with an option to add a global offset L_r which controls the trade-off between more saturated colors or less noise.

The proposed visualization scheme can be used alone, or in conjunction with existing tone mapping operators for HDR scenes. In this paper, we use the tone mapping operator proposed in [Mantiuk et al. 2006], but any existing operator can be applied. Fig. 9 shows the visualization result for different scene brightness. The default recommended parameter ($L_r = 0$) is highlighted in red, but users can make their own artistic choice by adjusting L_r as they like.

Relation to mesopic vision. The proposed visualization approach has parallels to the functioning of the human eye. Human vision covers a wide dynamic range. At high illuminance levels, cones (RGB sensors) are active and contribute to color vision, which is called photopic vision. At low illuminance levels, rods (W sensors) are active and only monochrome vision is achieved, which is called scotopic vision. For illuminance levels between these two extremes, both cones and rods are active, which is called mesopic vision [Kirk and O'Brien 2011]. Interestingly, mesopic vision marks a smooth transition between color vision and monochrome vision: Our vision system blends the signals from cones and rods to create an image [Shin et al. 2004]. The darker the scene, the more contribution is given to the rods, and the less the color we see.

Simulation of mesopic vision has been studied in the graphics community, which focuses on recreate various visual effects during mesopic vision such as desaturation and hue shift (Purkinje shift) on a bright monitor [Jacobs et al. 2015]. Besides colors, effects such as loss of acuity [Jacobs et al. 2015] and lowered resolution of perceived disparity [Kellnhofer et al. 2014] have also been recreated. However, these works assume that a clean, noise-free input image is available as input, which is not always the case when taking photos in dark environments. Although taking inspiration from these works, our main goal is to make the image appears less noisy. This is the reason why we only manipulate the color ratios to decouple measured high-SNR luminance channels and low-SNR chrominance channels, and we avoid complicated nonlinear operations that better model human vision but amplify noise.

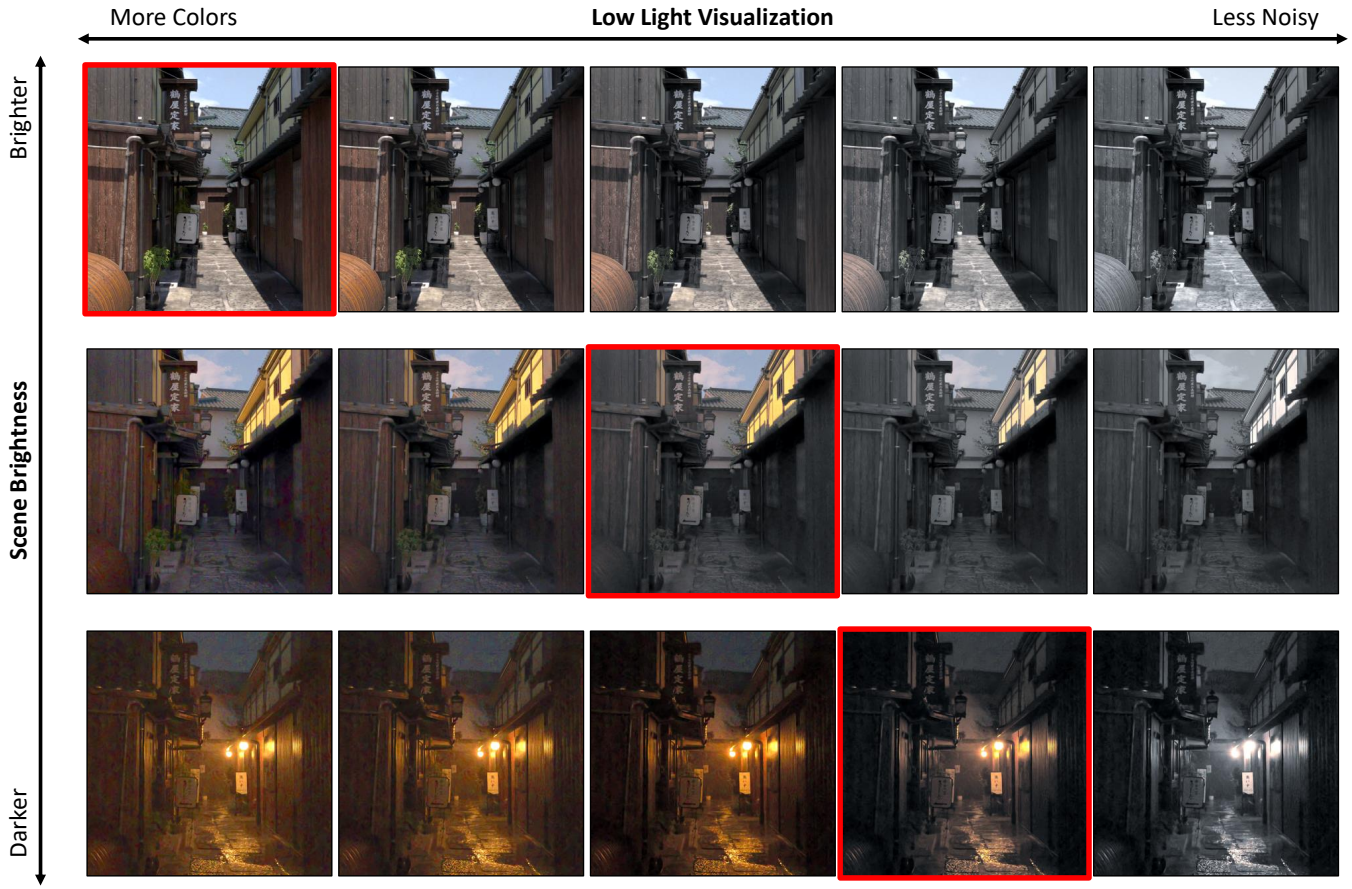

Fig. 9. **Noise-aware low-light visualization.** We show the visualization result for three different light levels on a simulated scene. The default visualization result is highlighted in red. In addition, we also provided users with an optional paramter L_r to control the trade-off between more saturated colors and less color noise, according to their aesthetic taste.

Can the proposed visualization help with RGB patterns? Recall that proposed method helps reduce noise because when an RGBW pattern is used, the luminance channel has a much higher SNR than the chrominance channels. An important ablation study question is: Will it help with images captured by an RGB pattern? Fig. 10 compares the visualization results of the same scene captured by RGBW (75% W) and RGB patterns. When we use a lower L_r (less noisy), the color noise in the RGBW result is suppressed, while the luminance noise in the RGB result remains, resulting in worse visual quality. Therefore, the proposed method is only helpful when an RGBW pattern is used.

User study. We conduct a user study to evaluate the efficacy of the proposed noise-aware visualization. We choose 30 photos taken in indoor or nighttime environments from the HDR+ dataset [Hasinoff et al. 2016] and synthesize a 2000-frame sequence of mosaicked binary images by linearly translating each image and then sampling from the Bernoulli distribution. Each pixel receives 0.002 photons per frame on average, and therefore the proposed align and merge algorithm generates a highly noisy image. The study consists of two rounds. During each round, the reconstructed noisy image, together with three mesopic mapped images with different global offset $L_r = -1, -0.5, 0$, are displayed simultaneously on a webpage. Both rounds consist of the same set of 30 photos, but the order of the photos and the order of the choices are randomized to avoid bias.

Fig. 11 shows the webpage shown to the participants in the study. A question is displayed at the top: Which image looks cleanest? (Round 1) / Which image do you like best? (Round 2). The four images are displayed at the same time so the participants can compare them side-by-side. The study is anonymous: Each participant visits the webpage from their own web browser, and we do not collect identity information such as name or IP address.

Fig 12 shows the result based on 900 votes from 15 participants. Q1 shows that the number of votes increases as more aggressive mapping is applied (less color), which validates our assumption that decreasing the weight of chrominance channels reduces chrominance noise and makes images perceptually cleaner. Q2 shows that the cleanest image is not always the one people prefer, due to the

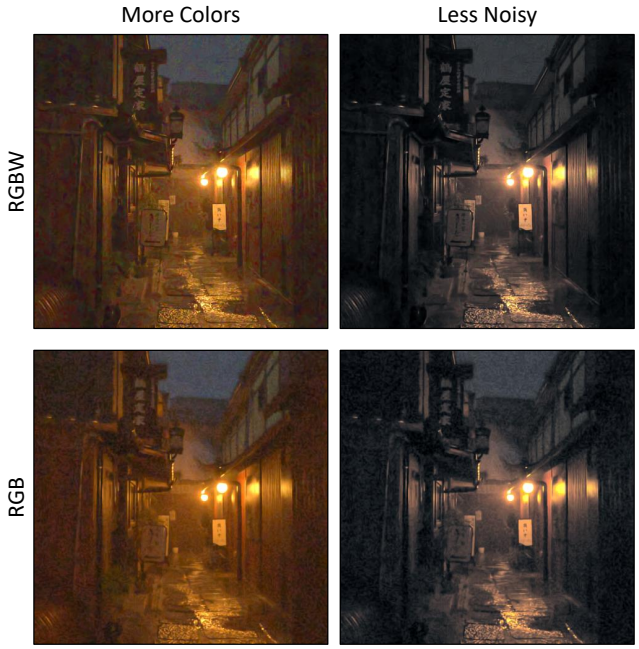

Fig. 10. **Noise-aware visualization for RGB filter array. (Left)** When more colors is preferred in the visualization, the RGBW pattern creates an image with more color noise, while the RGBW pattern creates an image with more luminance noise. **(Right)** When less noisy is preferred in the visualization, color noise in the RGBW image is suppressed, while the luminance noise in the RGB image remains, resulting in an image with lower visual quality (blurrier text while having significantly more noise). This shows that the proposed visualization scheme only helps with RGBW patterns.

color-noise trade-off. Most people prefer the default recommendation by the algorithm ($L_r = 0$).

REFERENCES

F J Anscombe. 1948. The Transformation of Poisson, Binomial and Negative-Binomial Data. *Biometrika* 35, 3/4 (1948), 246–254.

Ivan Michel Antolovic, Samuel Burri, Claudio Bruschini, Ron Hoebe, and Edoardo Charbon. 2016. Nonuniformity Analysis of a 65-Kpixel CMOS SPAD Imager. *IEEE Transactions on Electron Devices* 63, 1 (Jan. 2016), 57–64. https://doi.org/10.1109/TED.2015.2458295

Chenyan Bai, Jia Li, Zhouchen Lin, and Jian Yu. 2016. Automatic Design of Color Filter Arrays in The Frequency Domain. *IEEE Transactions on Image Processing* 25, 4 (2016), 1793–1807. https://doi.org/10.1109/TIP.2016.2531287

Folkmar Bornemann and Tom März. 2007. Fast Image Inpainting Based on Coherence Transport. *Journal of Mathematical Imaging and Vision* 28, 3 (Oct. 2007), 259–278. https://doi.org/10.1007/s10851-007-0017-6

Stanley Chan, Omar Elgendy, and Xiran Wang. 2016. Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors. *Sensors* 16, 11 (Nov. 2016), 1961. https://doi.org/10.3390/s16111961

Laurent Condat. 2009. A Generic Variational Approach for Demosaicking from an Arbitrary Color Filter Array. In *IEEE International Conference on Image Processing (ICIP)*. IEEE, Cairo, Egypt, 1625–1628. https://doi.org/10.1109/ICIP.2009.5413388

Oliver Cossairt, Mohit Gupta, and Shree K. Nayar. 2013. When Does Computational Imaging Improve Performance? *IEEE Transactions on Image Processing* 22, 2 (Feb. 2013), 447–458. https://doi.org/10.1109/TIP.2012.2216538

Omar A. Elgendy and Stanley H. Chan. 2020. Color Filter Arrays for Quanta Image Sensors. *IEEE Transactions on Computational Imaging* 6 (2020), 652–665. https://doi.org/10.1109/TCI.2020.2964238 arXiv:1903.09823

Edward B Gindele and Andrew C Gallagher. 2002. Sparsely sampled image sensing device with color and luminance photosites. US Patent 6,476,865.

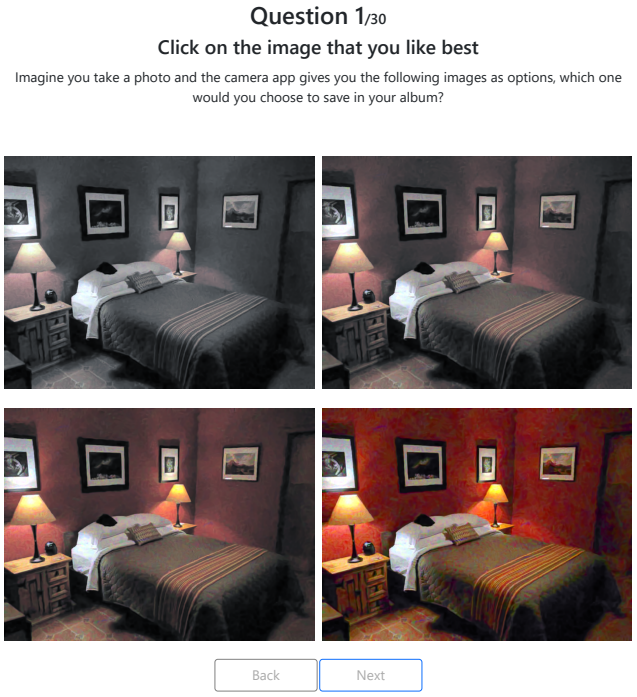

Fig. 11. **Screenshot of the user study webpage.**

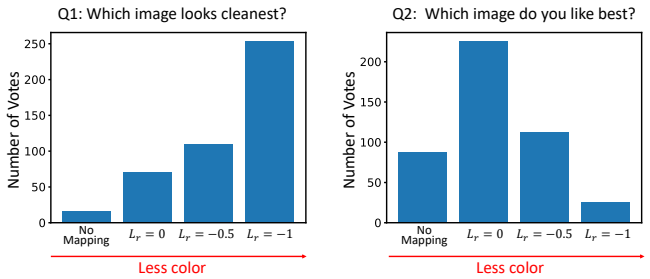

Fig. 12. **User study. (Left)** Most participants agree that aggressive mapping makes the images perceptually cleaner. **(Right)** Most participants do not favor images with least noise and prefer results with recommended setting.

Abhiram Gnanasambandam, Omar Elgendy, Jiaju Ma, and Stanley H. Chan. 2019. Megapixel Photon-Counting Color Imaging Using Quanta Image Sensor. *Optics Express* 27, 12 (June 2019), 17298. https://doi.org/10.1364/OE.27.017298

Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. 2016. Burst Photography for High Dynamic Range and Low-Light Imaging on Mobile Cameras. *ACM Transactions on Graphics* 35, 6 (Nov. 2016), 1–12. https://doi.org/10.1145/2980179.2980254

Bernardo Henz, Eduardo S. L. Gastal, and Manuel M. Oliveira. 2018. Deep Joint Design of Color Filter Arrays and Demosaicing. *Computer Graphics Forum* 37, 2 (May 2018), 389–399. https://doi.org/10.1111/cgf.13370

K. Hirakawa and P.J. Wolfe. 2008. Spatio-Spectral Color Filter Array Design for Optimal Image Recovery. *IEEE Transactions on Image Processing* 17, 10 (Oct. 2008), 1876–1890. https://doi.org/10.1109/TIP.2008.2002164

Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. 2022. Real-Time Intermediate Flow Estimation for Video Frame Interpolation. In *European Conference on Computer Vision (ECCV)*. Springer, 624–642. arXiv:2011.06294 [cs]

ITU-R. 2011. Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios. (2011).

David E. Jacobs, Orazio Gallo, Emily A. Cooper, Kari Pulli, and Marc Levoy. 2015. Simulating the Visual Experience of Very Bright and Very Dark Scenes. *ACM Transactions on Graphics* 34, 3 (May 2015), 1–15. https://doi.org/10.1145/2714573

Petr Kellnhofer, Tobias Ritschel, Peter Vangorp, Karol Myszkowski, and Hans-Peter Seidel. 2014. Stereo Day-for-Night: Retargeting Disparity for Scotopic Vision. *ACM Transactions on Applied Perception* 11, 3 (Oct. 2014), 1–17. https://doi.org/10.1145/2644813

Adam G Kirk and James F O'Brien. 2011. Perceptually Based Tone Mapping for Low-Light Conditions. *ACM Transactions on Graphics* 30, 4 (2011), 10.

Sizhuo Ma, Shantanu Gupta, Arin C. Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. 2020. Quanta Burst Photography. *ACM Transactions on Graphics* 39, 4 (July 2020), 1–16. https://doi.org/10.1145/3386569.3392470

Rafal Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2006. A Perceptual Framework for Contrast Processing of High Dynamic Range Images. *ACM Transactions on Applied Perception* 3, 3 (July 2006), 286–308.

S. M. A. Sharif and Yong Ju Jung. 2019. Deep Color Reconstruction for a Sparse Color Sensor. *Optics Express* 27, 17 (Aug. 2019), 23661. https://doi.org/10.1364/OE.27.023661

JaeChul Shin, Naoki Matsuki, Hirohisa Yaguchi, and Satoshi Shioiri. 2004. A Color Appearance Model Applicable in Mesopic Vision. *Optical Review* 11, 4 (July 2004), 272–278. https://doi.org/10.1007/s10043-004-0272-3

Hyeonjun Sim, Jihyong Oh, and Munchurl Kim. 2021. XVFI: eXtreme Video Frame Interpolation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. 14489–14498.